

Grado en Estadística y empresa  
2022-2023

*Trabajo Fin de Grado*

“Evolución de un indicador global creado sobre  
el ODS Agua Limpia y Saneamiento en las  
Comunidades Autónomas de España en los  
años 2016, 2018 y 2020”

---

Autor: Miguel Díaz-Plaza Cabrera

Tutora: Aurea Grané Chávez

Madrid, Julio 2023





## ÍNDICE DE CONTENIDOS

PRÓLOGO.....	II
RESUMEN.....	III
ÍNDICE DE FIGURAS.....	IV
ÍNDICE DE TABLAS.....	V
<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>2. BASE DE DATOS .....</b>	<b>4</b>
2.1. Descripción Base de Datos .....	4
2.2. Análisis descriptivo univariante .....	7
2.2.1. Variable Población .....	7
2.2.2. Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo .....	7
2.2.3. Coste unitario del agua .....	8
2.2.4. Volúmenes de agua registrada y distribuida a los usuarios .....	9
2.3. Análisis descriptivo multivariante y transformaciones necesarias. ....	10
<b>3. DESCRIPCIÓN DE LAS TÉCNICAS A UTILIZAR .....</b>	<b>12</b>
3.1. Análisis de Componentes Principales (PCA) .....	12
3.2. Construcción del Indicador .....	14
<b>4. DESARROLLO DEL ESTUDIO .....</b>	<b>15</b>
4.1. Construcción de Subindicadores por grupos temáticos .....	15
4.1.1. PCA: grupo Suministro.....	17
4.1.2. PCA: grupo Alcantarillado y depuración .....	22
4.2. Construcción del Indicador Global.....	25
<b>5. ANÁLISIS DE LOS RESULTADOS.....</b>	<b>30</b>
5.1. Visualización del Indicador Global .....	30
5.2. Relación Indicador Global con Variables Demográficas y Socioeconómicas .....	32
<b>6. CONCLUSIONES .....</b>	<b>34</b>
<b>7. BIBLIOGRAFÍA.....</b>	<b>36</b>
<b>8. ANEXO .....</b>	<b>37</b>

## **PRÓLOGO**

Este Trabajo de Fin de Grado surge como fruto de la subestimación, a concepción del autor, de uno de los recursos vitales indispensables como es el agua, y tiene como objetivo secundario concienciar al lector sobre su uso sostenible y responsable.

Mediante este trabajo se han procurado plasmar todos los conocimientos posibles adquiridos durante los cuatro años del grado de Estadística y empresa. Por este motivo, quisiera agradecer a todos los profesores y compañeros que me han acompañado en este precioso camino, en especial a la tutora de este trabajo, Aurea Grané Chávez, por enseñarme la utilidad de la estadística para solucionar problemas reales.

## RESUMEN

El ODS Agua limpia y saneamiento es el sexto de los 17 objetivos de desarrollo sostenible de la Agenda 2030 elaborada por Naciones Unidas. Tiene como meta garantizar el acceso universal al agua y saneamiento para el año 2030, y se trata de una necesidad vital para que se cumplan el resto de los objetivos. En este Trabajo de Fin de Grado se realizará un análisis de la situación de este ODS para cada una de las CCAA de España en un marco temporal de tres años: 2016, 2018 y 2020. Se efectuará mediante la creación de un indicador global, que asignará una valoración a cada CCAA en cada correspondiente año, por medio de la técnica estadística de análisis de componentes principales (PCA). Las variables participantes en el estudio han sido extraídas del INE (Instituto Nacional de Estadística), y resumen los indicadores de calidad propuestos por la ONU para el seguimiento de cumplimiento de este objetivo.

El primer paso, previo a la aplicación de la técnica estadística, es la depuración de los datos. Para ello se trasformarán las variables de forma que sean simétricas y sigan una distribución normal, y se realizará un análisis univariante y multivariante para identificar relaciones entre las variables.

El desarrollo del estudio se completará en dos fases. En una primera fase, se agruparán las variables en dos grupos temáticos, aplicando el PCA por separado para cada uno de los grupos y obteniendo un subindicador para cada grupo ponderando las componentes principales según su varianza explicada. Posteriormente, en la segunda fase, se repetirá el análisis de componentes principales sobre los subindicadores creados anteriormente, con el fin de lograr un indicador global que resuma la información total. De modo que, el resultado estará compuesto por tres valores del indicador global para cada CCAA, uno correspondiente a cada año.

Finalmente, el trabajo concluirá con la interpretación de estos valores del indicador global, estudiando la comparativa del mismo a lo largo del espacio temporal (tanto numéricamente como de forma gráfica), clasificando las distintas CCAA's en grupos, y analizando la relación de este indicador global con otras variables socioeconómicas, como la población o el IPC.

## Palabras clave

Agua; CCAA; componente principal; España; grupos temáticos; indicador; ODS; PCA

## ÍNDICE DE FIGURAS

Figura 1. Cobertura mundial de servicios de agua potable, saneamiento e higiene, con datos reales entre 2015-2020 (porcentaje) y el incremento necesario para alcanzar la cobertura universal para 2030.

Figura 2. Nivel de estrés hídrico: extracción de agua dulce en proporción al total de recursos renovables de agua dulce, 2019 (porcentaje).

Figura 3. Comunidades autónomas con mayor población promedio entre los años 2016, 2018 y 2020.

Figura 4. Mapa coroplético de Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo en el año 2016, 2018 y 2020.

Figura 5. Volumen de agua registrada y distribuida en función del tipo de usuario por CCAA, en promedio de todos los años del estudio (2016-2018-2020).

Figura 6. Autovalores del PCA del grupo Suministro

Figura 7. Interpretación gráfica de las dos primeras componentes del grupo Suministro

Figura 8. Autovalores del PCA del grupo Alcantarillado y depuración

Figura 9. Interpretación gráfica de las dos primeras componentes del grupo Alcantarillado y depuración

Figura 10. Interpretación gráfica de las dos primeras componentes del nuevo conjunto de datos reducido

Figura 11. Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2016.

Figura 12. Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2018.

Figura 13. Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2020.

Figura 14. Comparativa boxplots del Indicador global entre los años 2016, 2018 y 2020.

Figura 15. Diagrama de dispersión matricial con su respectiva matriz de correlaciones, entre las variables X1 y X2 transformadas, el indicador global, y los subindicadores temáticos de suministro y de alcantarillado y depuración

## ÍNDICE DE TABLAS

Tabla 1. Representación de todas las variables y su correspondiente unidad de medida.

Tabla 2. Media del coste unitario del agua, dividido en coste unitario de suministro y coste unitario del alcantarillado y depuración en cada año del estudio (2016-2018-2020).

Tabla 3. Subgrupos y variables que pertenecen a cada subgrupo.

Tabla 4. Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del grupo Suministro

Tabla 5. Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del grupo Alcantarillado y depuración

Tabla 6. Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del nuevo conjunto de datos

Tabla 7. Autovectores de cada componente principal del nuevo conjunto de datos reducido

Tabla 8. Puntuaciones del Indicador global por CCAA y por año

# 1. INTRODUCCIÓN

El acceso al agua limpia y saneamiento es el derecho humano más básico y necesario para sobrevivir, y es el que sostiene el resto de Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 de la Organización de Naciones Unidas.

Los Objetivos de Desarrollo Sostenible (ODS) se tratan de 17 metas propuestas y aprobadas por las Naciones Unidas en 2015, que fueron creadas con el fin de mejorar el bienestar de la población mundial, además de fines consecuentes como proteger la sostenibilidad del planeta o acabar con la pobreza. Estos aspectos conforman el plan Agenda 2030 de la ONU, con el propósito de evaluar en 15 años la situación de cada país respecto al desempeño en estos objetivos. (Naciones Unidas, 2023)

Cada ODS abarca un tema de la realidad diferente, estando todos muy relacionados entre sí. Concretamente, el ODS a tratar es el número 6, que pretende un acceso universal al agua limpia y saneamiento, es vital para que se cumplan el resto de las metas. (Naciones Unidas, 2023)

El agua es un elemento vital para la salud humana, sirve tanto para hidratarse como para mantener una higiene adecuada, al mismo tiempo que permite llevar a cabo nuestras funciones diarias y prevenir enfermedades. Un acontecimiento histórico relevante, que ha realzado la importancia del acceso universal al agua potable, el saneamiento y la higiene, es el COVID-19. Una correcta higiene de manos con agua y jabón es básico y fundamental para protegerse del Coronavirus, así como de muchas otras enfermedades. De ahí que, la forma de gestionar y distribuir el agua limpia puede evitar una gran cantidad de enfermedades y muertes. En definitiva, la pandemia del COVID-19 nos ha hecho darnos cuenta de la importancia vital de estos servicios. (Naciones Unidas, 2023)

Tal y como refleja el Informe de los Objetivos de Desarrollo Sostenible 2022 de las Naciones Unidas, durante los últimos años, el problema del acceso al agua limpia y saneamiento se ha ido expandiendo tanto que, aunque en los esfuerzos año tras año se observa un progreso, las cifras siguen siendo alarmantes. Según manifiesta el informe, se tendría que cuadruplicar el ritmo de mejora para alcanzar las metas de la Agenda 2030. Lograr este objetivo permitiría salvar 829000 vidas al año, que son las personas que mueren al año directamente por agua insalubre, saneamiento inadecuado y malas prácticas de higiene. (Informe de los Objetivos de Desarrollo Sostenible, 2022).

Los servicios de agua potable gestionados de manera segura aumentaron del 70% del 2015 al 74% en 2020, lo que supone aun así que 2000 millones de personas carecieran de estos servicios. Al ritmo actual de progreso, en el 2030 se podría alcanzar el 81%, por lo que no se cumpliría la meta y se quedarían todavía 1600 millones de personas sin suministro de agua potable. (Informe de los Objetivos de Desarrollo Sostenible, 2022).

En cuanto a los servicios de saneamiento gestionados de manera segura, durante el periodo de 2015 al 2020, aumentaron del 47% al 54% y siguiendo el ritmo histórico de progreso

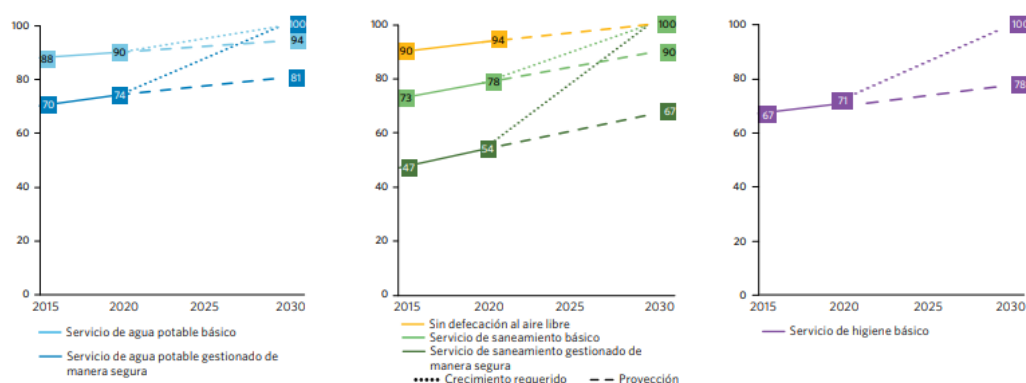


alcanzaríamos el 67% para el año 2023, dejando a 2800 millones de personas sin estos servicios. (Informe de los Objetivos de Desarrollo Sostenible, 2022).

Respecto a las prácticas de higiene, la defecación al aire libre se redujo en un tercio, de 739 millones de personas a 494 millones y se prevé que para el 2030 esta práctica se eliminará. La higiene de manos es primordial para controlar las enfermedades infecciosas, como hemos visto con el COVID-19, sin embargo 1 de 4 personas no tienen acceso a instalaciones de agua para lavarse las manos con agua y jabón. (Informe de los Objetivos de Desarrollo Sostenible, 2022).

**Figura 1**

*Cobertura mundial de servicios de agua potable, saneamiento e higiene, con datos reales entre 2015-2020 (porcentaje) y el incremento necesario para alcanzar la cobertura universal para 2030.*



*Fuente: Informe de los Objetivos de Desarrollo Sostenible, 2022*

Es evidente que el crecimiento acelerado de la población mundial es uno de los factores que ponen en riesgo la distribución del agua, ya que supone una mayor demanda y un mayor agotamiento de los recursos.

Otro de los factores que contribuye a ello, y que será de gran interés en este estudio, es la agricultura, la industria y el sector energético. Según datos de Naciones Unidas, más del 80% de las aguas residuales consecuentes de la acción humana a nivel mundial son vertidas en ríos y mares sin tratamiento previo. Una gestión más sostenible de los recursos hídricos, como un mejor tratamiento de las aguas residuales, podría impedir que se dañaran los ecosistemas. En adicción, los humedales del mundo se están perdiendo a un ritmo frenético, y los ríos, lagos o embalses están sufriendo alteraciones drásticas, por lo que es indispensable proteger estos hábitats biológicos. (Naciones Unidas, 2023).

A pesar de que la evolución tiene tendencia alcista, los datos a nivel mundial nos otorgan una considerable preocupación. De hecho, la meta del agua limpia y saneamiento es el ODS del mes de marzo de 2023. Conjuntamente, es importante concienciar a la población, visibilizando el problema en el Día Mundial del Agua (se celebró el 22 de marzo de 2023) y con iniciativas que se están poniendo en práctica en universidades, escuelas u oficinas de trabajo, que consisten en colocar un punto dispensador de agua para que todo el mundo tenga acceso libre.

En marzo de 2023, se celebró la Conferencia del Agua 2023 de la ONU en Nueva York, en la que se trataron de impulsar determinadas acciones para conseguir aquellas metas relacionadas con el agua que se están viendo retrasadas para la Agenda 2030, y, cuyo incumplimiento se puede ver reflejado en el resto de los objetivos. Y es que marzo de 2023 es un mes clave para el objetivo de agua limpia y saneamiento, ya que el 20 de marzo, el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) publicó su Sexto Informe de Evaluación, un informe sobre las medidas del cambio climático que se publica cada 6 o 7 años. Además, el 21 de marzo, la UNESCO lanzó el Informe Mundial sobre el Desarrollo de los Recursos Hídricos (WWDR) de la ONU-Agua que, como cada año, propone mejoras y un profundo análisis de las cuestiones más recientes. (Conferencia ONU 2023 sobre el Agua, marzo 2023, Nueva York)

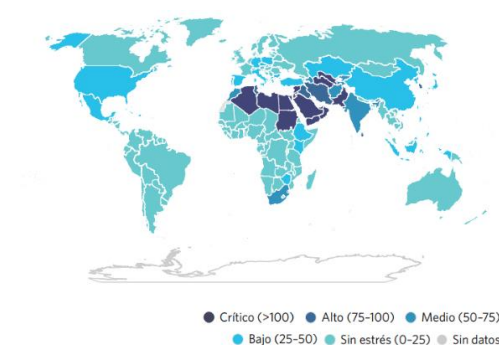
Sin embargo, los problemas relacionados con el agua y el saneamiento tienen una estrecha relación con la situación socioeconómica y el desarrollo de cada país.

Aunque en 2019, el estrés hídrico mundial alcanzó un promedio del 18,6%, existen zonas del África septentrional y Asia occidental que sufrieron un aumento del 13% desde el 2015, alcanzando un 84,1%. Aproximadamente, el 10% de la población mundial reside en países con un nivel de estrés hídrico alto y crítico, cifra equivalente a 733 millones de personas. La solución para apaciguar este estrés es mejorar la eficiencia en el uso del agua en estos países subdesarrollados. (Informe de los Objetivos de Desarrollo Sostenible, 2022).

Una vía para alcanzar esta solución es establecer acuerdos de cooperación entre países que comparten ríos, lagos o acuíferos, para garantizar la gestión equitativa y sostenible de estos recursos

## Figura 2

*Nivel de estrés hídrico: extracción de agua dulce en proporción al total de recursos renovables de agua dulce, 2019 (porcentaje)*



*Fuente: Informe de los Objetivos de Desarrollo Sostenible, 2022*

En este trabajo, se analiza la evolución en el territorio español del mencionado sexto Objetivo de Desarrollo Sostenible de la Agenda 2030, que mide la disponibilidad de agua y su gestión sostenible y saneamiento para cada comunidad autónoma perteneciente al Estado español. Dicho análisis comparativo de la evolución se estudiará en un marco temporal de tres años: 2016, 2018 y 2020.

## 2. BASE DE DATOS

### 2.1. Descripción base de datos.

La base de datos construida está compuesta por 22 variables sobre 18 Comunidades Autónomas del territorio español, abarcando el espacio temporal de entre los años 2016, 2018 y 2020. Es decir, dispondremos de 3 tablas, una correspondiente al año 2016, otra al 2018 y por último al 2020.

Las Comunidades Autónomas sobre las que se basará el estudio serán las 17 comunidades autónomas de España, más Ceuta y Melilla, las dos ciudades con estatuto de autonomía del Estado español, que fueron unificadas en una sola comunidad debido a que se disponía de indicadores que las otorgaban datos por separado y otros que las trataban de forma conjunta.

De las 22 variables que se analizarán en este estudio, únicamente una de ellas corresponde íntegramente a un indicador del Objetivo de Desarrollo Sostenible número 6, agua limpia y saneamiento. Está incluido en la meta 6.1., que consiste en *De aquí a 2030, lograr el acceso universal y equitativo al agua potable a un precio asequible para todos.* y concretamente se trata del indicador 6.1.1, que mide la *Proporción de la población que utiliza servicios de suministro de agua potable gestionados sin riesgos.* (Naciones Unidas, 2023)

De las variables restantes, 19 de ellas se tratan de variables relacionadas con el agua y el saneamiento, que se intentan aproximar en la mayor medida de lo posible a lo que miden el resto de los indicadores de los que no se disponen datos para el espacio temporal y territorios deseados.

Por otra parte, las otras dos variables, son variables sociodemográficas como son la población de cada una de las Comunidades Autónomas en cada año correspondiente; y el IPC (índice de Precios de Consumo) correspondiente al subgrupo 044 Suministro de agua y otros servicios relacionados con la vivienda sobre el año base 2021.

Todas las variables han sido extraídas del Instituto Nacional de Estadística. (INE - Instituto Nacional de Estadística, s. f.)

A continuación, se muestra en la *tabla 1*, una representación de todas las variables que se disponen para el estudio, con su correspondiente unidad de medida.

**Tabla 1***Representación de todas las variables y su correspondiente unidad de medida.*

<b>Variable</b>	<b>Descripción</b>	<b>Unidad de medida</b>
<b>X1</b>	Población	Unidades
<b>X2</b>	IPC Suministro de agua y otros servicios relacionados con la vivienda	Índice BASE 2021
<b>X3</b>	Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo (Indicador 6.1.1)	%
<b>X4</b>	Coste Unitario Agua	€/ m <sup>3</sup>
<b>X5</b>	Coste Suministro	€/ m <sup>3</sup>
<b>X6</b>	Coste Alcantarillado y depuración	€/ m <sup>3</sup>
<b>X7</b>	Importe facturado por el agua suministrada	Miles de €
<b>X8</b>	Importe facturado por alcantarillado y depuración	Miles de €
<b>X9</b>	Longitud red de suministro por habitante	m/habitante
<b>X10</b>	Longitud red de alcantarillado por habitante	m/habitante
<b>X11</b>	Pérdidas reales por km de suministro	m <sup>3</sup> /km
<b>X12</b>	Importe total de la inversión en servicios de suministro	Miles de €
<b>X13</b>	Número de tanques de tormenta	Unidades
<b>X14</b>	Volumen total de agua registrada y distribuida en sectores económicos	Miles m <sup>3</sup>
<b>X15</b>	Volumen total de agua registrada y distribuida en hogares	Miles m <sup>3</sup>
<b>X16</b>	Volumen total de agua registrada y distribuida en consumos municipales	Miles m <sup>3</sup>
<b>X17</b>	Porcentaje pérdidas reales sobre el volumen de agua suministrada	%
<b>X18</b>	Volumen de agua disponible no potabilizada	Miles m <sup>3</sup>
<b>X19</b>	Volumen de agua disponible para su potabilización	Miles m <sup>3</sup>
<b>X20</b>	Volumen de aguas residuales tratadas	m <sup>3</sup> /día
<b>X21</b>	Volumen total de agua reutilizada	m <sup>3</sup> /día
<b>X22</b>	Volumen de lodos generados en el tratamiento de aguas residuales	Toneladas de materia seca

La variable X1, correspondiente a la población, consta de cifras oficiales resultantes de la revisión del Padrón municipal a 1 de enero según el INE. La población de Ceuta y Melilla ha sido calculada mediante la suma de la población respectiva a Ceuta más la referente a Melilla.

Para Ceuta y Melilla, las variables X2 y X3, que son el IPC Suministro de agua y otros servicios relacionados con la vivienda y el porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo, respectivamente, han sido calculadas mediante una media ponderada según la población de cada provincia.

X2 al ser un índice tiene de base el año 2021, lo que quiere decir que ese año posee el valor 100, y el resto de los valores se interpretan en función de ese valor. Por su parte, X3 se mide en porcentajes, oscilando los valores entre 0% y 100%.

La variable X4, coste unitario del agua, es la suma de las consecutivas variables X5 y X6, coste suministro y coste alcantarillado y depuración, respectivamente. Este trío de variables se mide en €/m<sup>3</sup>, o lo que es lo mismo, €/1000 litros de agua.

Siguiendo con las variables que explican datos económicos, tenemos X7 y X8, que son los importes facturados por el agua suministrada y por alcantarillado y depuración, respectivamente, medidas en miles de €.

Continuando con la comparativa entre red de suministro y red de alcantarillado y depuración, tenemos en las variables X9 y X10 su respectiva longitud, medida en m/habitante. Este es un campo recalculado, ya que previamente se disponía de la longitud en km, y cobraba más sentido expresarlo en función de la población y en metros.

Relacionado con los servicios de suministro, hemos añadido otras tres variables, como son X11, X12 y X13. X11 representa las pérdidas reales por km en la red de suministro, que se mide en m<sup>3</sup>/km; X12 muestra el importe total de la inversión en servicios de suministro, medido en miles de €; y X13 es el número de tanques de tormenta.

Cambiando a otro bloque de variables, tenemos X14, X15 y X16, que representan el volumen total de agua registrada y distribuida por usuario. Siendo X14 en sectores económicos, X15 en hogares y X16 en consumos municipales. Todas medidas en miles de m<sup>3</sup>. Otra variable que está relacionada con las anteriores es X17, que indica el porcentaje de pérdidas reales sobre el volumen de agua suministrada.

Las variables X18 y X19 muestran el volumen de agua disponible no potabilizada y potabilizada, respectivamente, en miles de m<sup>3</sup>.

Por último, las variables X20, X21 y X22 están relacionadas con las aguas residuales. X20 representa el volumen de aguas residuales tratadas y X21 el volumen total de agua reutilizada, ambas medidas en m<sup>3</sup>/día. Para finalizar, X22 es una medida del volumen de lodos generados en el tratamiento de aguas residuales proporcionada en toneladas de materia seca.

## 2.2. Análisis descriptivo univariante.

En este apartado, se realizará un análisis descriptivo detallado de las variables consideradas a priori más relevantes y que pueden influir notablemente a los resultados de este estudio. Para ello, se aportará una forma de representación gráfica que permita visualizar datos significativos que se puedan destacar, o en caso de que sea necesario, medidas estadísticas que aporten valor a la información.

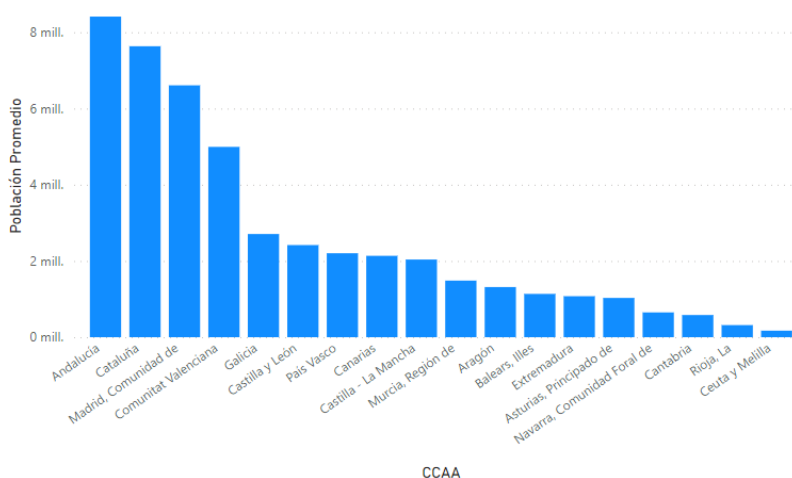
En la *figura A.1.*, que se encuentra en el anexo, se ofrece un diagrama de caja de cada variable en particular.

### 2.2.1. Variable Población.

Se comenzará explicando la variable X1, correspondiente a la población. Se trata de una variable demográfica, externa a los objetivos de desarrollo sostenible, incluida para contextualizar y enriquecer la base de datos. Como se observa en la siguiente *figura 3*, la comunidad autónoma con mayor población de media en los años 2016, 2018 y 2020 se trata de Andalucía, superando los 8 millones de habitantes.

**Figura 3**

*Comunidades autónomas con mayor población promedio entre los años 2016, 2018 y 2020.*



*Fuente: Microsoft Power BI*

### 2.2.2. Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo.

La variable X3 corresponde a la meta 6.1. de los ODS y se trata de lograr el acceso universal y equitativo a agua potable segura y asequible para todos antes de 2030.

En concreto, se trata del Indicador 6.1.1, que indica el porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo.

En las siguientes figuras, se concluirá si se observa evolución en este objetivo a medida que aumentan los años de estudio, a través de tres mapas coropléticos, uno correspondiente a cada año, en la que se pueden observar aquellas comunidades con un porcentaje menor en un tono más rojo y las que tienen un mayor porcentaje en color más verde, indicando que es más positivo ese valor para el cumplimiento del objetivo.

En la *figura 4*, del año 2016, la comunidad autónoma con menor porcentaje es la Región de Murcia con un 60,20%, mientras que la que mayor porcentaje tiene es, con un 100%, Castilla y León. Es decir, en 2016 Castilla y León consiguió la meta óptima que se quiere alcanzar, que toda la población tenga acceso a agua potable, se verá si lo puede mantener hasta 2030.

De nuevo en la *figura 4* correspondiente al año 2018, se contempla como la Región de Murcia obtiene el porcentaje más bajo con un 55,28% y Castilla y León alcanza prácticamente la perfección con un 99,99%.

En la *figura 4* correspondiente a 2020, se palpa más la igualdad entre todas las comunidades. Castilla y León sigue siendo la primera en el ranking con 98,58%, mientras que la más atrasada deja de ser Región de Murcia, que tiene un crecimiento significativo en 2020, y pasa a ser Illes Balears, con un porcentaje de 60,10%.

#### Figura 4

*Mapa coroplético de Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo en los años 2016, 2018 y 2020, de izquierda a derecha.*



*Fuente: Microsoft Power BI*

En general, los porcentajes son bastante positivos, pero no suficientes para alcanzar la meta de la Agenda 2030 de un acceso universal al agua potable, que supone conseguir el 100%.

#### 2.2.3. Coste unitario del agua.

El coste unitario del agua se define como el cociente entre el importe facturado por el agua suministrada (variable X7) junto al importe facturado por alcantarillado, depuración y todas las actividades relacionadas con el saneamiento (variable X8), y el volumen total de agua

registrada y distribuida a todos los usuarios (suma de las variables X14, X15, X16). (Estadística sobre el Suministro y Saneamiento del Agua, Año 2020).

Por tanto, este coste unitario del agua, que corresponde a la variable X4, se subdivide en otras dos variables, X5 y X6, que son respectivamente el coste unitario del suministro y el coste unitario del alcantarillado y depuración.

**Tabla 2**

*Media del coste unitario del agua, dividido en coste unitario de suministro y coste unitario del alcantarillado y depuración en cada año del estudio (2016-2018-2020).*

*(Unidad: €/m<sup>3</sup>)*

	Año 2016	Año 2018	Año 2020
Coste unitario de suministro	1,071	1,048	1,048
Coste unitario del alcantarillado y depuración	0,7033	0,7167	0,725
<b>Coste unitario del agua</b>	<b>1,774</b>	<b>1,764</b>	<b>1,773</b>

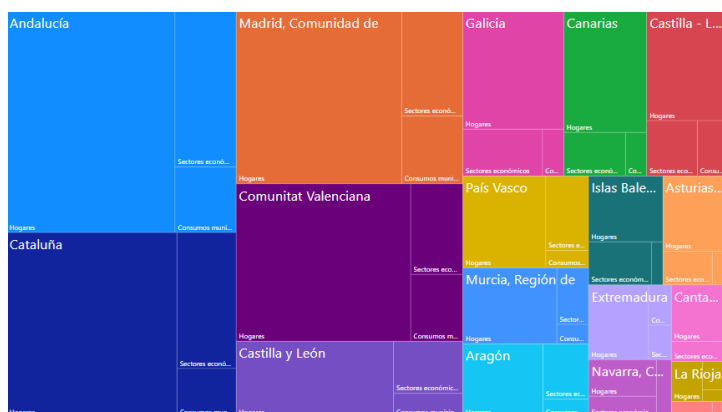
#### 2.2.4. Volúmenes de agua registrada y distribuida a los usuarios.

Los niveles de agua registrada y distribuida se pueden dividir según el tipo de usuario al que vaya dirigido. La variable X14 corresponde a los sectores económicos, X15 pertenece a los hogares y X16 a los consumos municipales.

En la *figura 7*, se puede apreciar que, del total de agua registrada y distribuida a los usuarios, las comunidades autónomas que más agua reciben en promedio de todos los años del estudio son Andalucía, Cataluña o la Comunidad de Madrid, y en prácticamente la mayoría de comunidades se cumple que el sector que más recibe son los hogares, seguidos de los sectores económicos y los consumos municipales.

**Figura 5**

*Volumen de agua registrada y distribuida en función del tipo de usuario por CCAA, en promedio de todos los años del estudio (2016-2018-2020)*



Fuente: Microsoft Power BI



### 2.3. Análisis descriptivo multivariante y transformaciones necesarias.

En este apartado se aplicarán algunas de las técnicas de análisis multivariante más importantes con el objetivo de conocer la relación que tienen las variables de la base de datos entre sí. Estas serán: vector de medias, matriz de correlaciones representado gráficamente y diagrama de dispersión matricial.

De esta manera, se intentará detectar asimetría en las variables y se llevará a cabo una transformación de estas, si es considerado necesario, para posteriormente aplicar las técnicas estadísticas. El objetivo será conseguir normalidad en cada una de las variables, o en caso contrario que no se pueda conseguir, buscar la simetría de los datos.

En el anexo se puede encontrar el vector de medias de las variables sin transformar por cada instante de tiempo en la *tabla A.1*.

En la *figura A.2*. del anexo también se puede visualizar la matriz de correlaciones del que se extraen conclusiones llamativas. Existen muchas correlaciones altas positivas que superan 0,80, como las de las variables X7, X8, X12, X14, X15, X16, X18, X19, X20 y X22 con respecto a X1 (Población), que indican que un aumento la población implica un aumento en estas variables. Es decir, se trata de variables que reflejan volúmenes totales y por tanto están altamente correlacionadas con la población. Cuanto mayor es el número de habitantes, mayor es el volumen que representan estas variables.

La solución para tratar estas variables será transformarlas dividiendo entre la población (X1) de forma que muestren un valor significativo que no dependa de dicha variable demográfica. En la *figura A.2.b* del apéndice se aprecia que ya no hay variables altamente correlacionadas con X1. Por tanto, a priori, no va a haber variables con una estrecha relación con la población de la CCAA.

El diagrama de dispersión matricial de todas las variables también se puede encontrar en el anexo en la *figura A.3.*, e indica la distribución de los datos, siendo el color rojo el año 2016, el 2018 el verde y el 2020 el azul.

Las variables que no parecen normales y presentan asimetría son: con asimetría izquierda: X1, X7, X8, X11, X12, X13, X14, X15, X16, X17, X18, X19, X20, X21, X22; y con asimetría derecha: X2, X3. Por tanto, las variables que parecen normales y no van a sufrir alteraciones son un total de 5 y son: X4, X5, X6, X9, X10.

Posteriormente, se aplica una transformación logarítmica a las variables por tratarse de asimetrías, excepto a X13 Y X21 por tener valores iguales a 0, que se diferencia en que se aplican logaritmos de las (variables +1). Es decir, como los logaritmos son invariantes a transformaciones y cambios de escala, se le suma una unidad al valor de todas las observaciones de dichas variables citadas en última instancia.

Las *figuras A.4.* (antes) y *A.5.* (después) que se muestran en el anexo evidencian que todas las variables parecen haber obtenido la normalidad, excepto X17 y X21.

X17 genera una asimetría a la derecha no deseable, así que se mantiene como estaba antes de la transformación.

X21 tiene muchos valores que son 0, si se eliminasen las observaciones que son 0 se reduciría el número de observaciones con dos comunidades menos. Por tanto, se procede a no tener en cuenta a esta variable porque es asimétrica, no va a conseguir la normalidad y va a ser perjudicial para la posterior aplicación de la técnica estadística.

Por su parte, como X4 está definida como la suma de X5 + X6, se va a descartar para no tener variables linealmente independientes que no vengan de los propios datos.

Tampoco se tendrá en cuenta a la hora de aplicar la correspondiente técnica estadística la variable población (X1) y la variable correspondiente al IPC del suministro de agua (X2), ya que se tratan de variables de carácter demográfico y socioeconómico que ofrecen información sobre el contexto y que no tendría sentido incluirlas en el estudio para construir el indicador global. Contrariamente, se utilizará esta pareja de variables para contrastar los resultados del indicador global, y ver qué relación poseen los valores del indicador respecto a estas variables de población e IPC y cómo evoluciona a lo largo de los años del estudio.

En conclusión, tras la aplicación de las transformaciones necesarias, se dispondrán de 18 variables para cada año de marco temporal (2016, 2018 y 2020) para el estudio de la técnica estadística: Análisis de Componentes Principales.

### 3. DESCRIPCIÓN DE LAS TÉCNICAS A UTILIZAR

#### 3.1. Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística de Análisis Multivariante que fue creada en 1891 por Karl Pearson, y mejorada posteriormente en 1933 por Harold Hotelling, con el objetivo principal de reducir la dimensión de conjuntos de datos con numerosas variables. En otras palabras, trata de describir la información comprendida en las variables observadas mediante un número menor de variables. Estas nuevas variables son las llamadas componentes principales. (Baíllo, A. and Grané, 2007)

Sea  $X = (X_1, \dots, X_p)'$  una matriz de datos centrada que representa el conjunto de datos originales de  $n$  filas (observaciones) y  $p$  columnas (variables originales), las componentes principales (CPs) se pueden describir como combinaciones lineales de las variables originales, que buscan explicar la mayor cantidad de varianza de  $X$ .

Estas componentes principales  $Z_1, Z_2, \dots, Z_p$  se ordenan descendentemente según la mayor cantidad de varianza explicada, y están incorreladas entre sí de la siguiente manera:

$$Z_1 = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p$$

$$Z_2 = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p$$

...

$$Z_p = a_{p1} X_1 + a_{p2} X_2 + \dots + a_{pp} X_p$$

Tomando como referencia la primera componente principal  $Z_1$ , e idénticamente para las demás  $Z_p$ , los términos  $a_{11}, \dots, a_{1p}$  son los pesos o loadings de la primera componente principal. Representan el peso o la importancia de cada variable en la componente. Es decir, cuanto mayor sea en valor absoluto el loading de una variable, más influye en la componente principal. Por tanto, ayudan a identificar qué tipo de información recoge cada variable y si influye de manera positiva o negativa a la componente principal. Cada combinación lineal es normalizada, lo que implica que la suma de todos los loadings al cuadrado de cada componente principal es igual a uno:

$$\sum_{j=1}^p a_{j1}^2 = 1$$

De tal forma, la primera componente principal  $Z_1$  será la que mayor porcentaje de varianza explicará de las variables originales. Habrá tantas componentes principales como variables originales se dispongan, de manera que el 100% de la variabilidad de las variables originales será el explicado por todas las componentes principales.

El problema del PCA reside en encontrar un número de componentes principales  $q$  menor que el de variables originales ( $q < p$ ), tal que sean capaces de maximizar la varianza explicada sobre la base de datos original. Por este motivo, el PCA es una técnica muy útil cuando se dispone de un elevado número de variables originales, por su capacidad para resumir la mayor cantidad de dicha información mediante un número menor de variables creadas.

Entre los criterios para decidir el subconjunto de componentes principales  $q < p$  a seleccionar, los tres más populares son:

- Scree-plot de Cattell: se basa en la representación gráfica de los autovalores  $\lambda_p$ , buscando la figura de un “codo”, punto a partir del cual los sucesivos valores propios son aproximadamente iguales y no se producen descensos bruscos de pendiente.
- Criterio de Kaiser: consiste en escoger aquellos componentes cuyos autovalores sean mayores que 1. La lógica de este criterio se fundamenta en que no tiene sentido escoger componentes principales que expliquen menos varianza que el valor de la varianza media de los componentes que es 1.
- Criterio de Jolliffe: es la modificación al criterio de Kaiser. Este criterio descarta aquellas componentes cuyos autovalores sean menores que 0,7, incluyendo más componentes que el criterio de Kaiser.

Concretamente en nuestro estudio, este último criterio de Jolliffe será el que se utilizará para determinar el número de componentes principales en cada caso.

Las componentes principales se obtienen mediante la diagonalización de la matriz de covarianzas de  $X$ , es decir, calculando sus autovalores y autovectores. La matriz de covarianzas ( $S$ ) tiene la siguiente forma:

$$S = T\Lambda T'$$

donde  $T$  es la matriz de autovectores. Es una matriz ortogonal  $p \times p$  cuyas columnas son los coeficientes de las Componentes Principales. Por su parte,  $\Lambda$  es la matriz diagonal de autovalores ordenados decrecientemente.

También se pueden conseguir a través de la matriz de correlaciones ( $R$ ), que es la matriz de covarianzas estandarizada:

$$R = \tilde{T}\tilde{\Lambda}\tilde{T}'$$

El cálculo de las componentes principales mediante la matriz de covarianzas o la matriz de correlaciones tiene una gran influencia en los resultados de las componentes, que distan bastante dependiendo de qué matriz se utilice. Para responder a la cuestión de si utilizar la matriz de covarianzas o correlaciones, necesitamos conocer las varianzas de las variables originales, así como sus unidades de medida.

Cuando una variable tiene una varianza más elevada que el resto, la primera componente principal coincidirá en gran medida con esta variable de varianza elevada y explicará una gran variabilidad del sistema. Así mismo pasa cuando una variable posee una unidad de medida mayor que el resto. Para estos casos, conviene utilizar la matriz de correlaciones, ya que estandariza los datos y realizar el análisis de componentes principales a partir de esta matriz de correlaciones  $R$ , en lugar de emplear la de covarianzas  $S$ . (Peña, 2002).

### 3.2. Construcción del Indicador

Una vez escogidas las correspondientes componentes principales que reducen la dimensión del conjunto de datos mediante los criterios mencionados anteriormente (se utilizará el criterio de Jolliffe en este estudio), el objetivo del PCA es construir un indicador.

Ese indicador se trata de un valor que simplifica a modo de puntuación, el resultado que obtiene cada CCAA en la aplicación del PCA. La fórmula que hay que aplicar para construir el indicador es la siguiente:

$$I = \sum_{i=1}^q \frac{\lambda_i}{\sum_{j=1}^q \lambda_j} \cdot Y_i$$

Donde  $q$  es el número de componentes principales escogidas y el cociente de lambdas es una ponderación de variabilidades explicadas, siendo el numerador  $\lambda_i$  la variabilidad explicada por componente  $i$  y el denominador  $\sum_{j=1}^q \lambda_j$  la variabilidad acumulada por las componentes  $q$ .  $Y_i$  se calcula multiplicando la matriz de datos original por cada autovector correspondiente.

Con el objetivo de que todos los indicadores estén en una misma escala comparable, se estandarizará el indicador obtenido en un rango de valores de 0 a 100, de la siguiente manera:

$$I' = \frac{I - \min(I)}{\max(I) - \min(I)} \cdot 100$$

Adquiriendo el valor  $I' = 0$  para el valor más bajo de  $I$ , y el valor  $I' = 100$  para el valor máximo de  $I$ , generando así un ranking con el que poder comparar las distintas CCAA.

(Grané and Albarrán, 2023)

## 4. DESARROLLO DEL ESTUDIO

### 4.1. Construcción de Subindicadores por grupos temáticos

Debido a que la base de datos originaria dispone de más predictores  $p$  que número de observaciones  $n$ , no se puede aplicar directamente el análisis de componentes principales. En este caso al tener 3 años de marco temporal, cada uno de los tres conjuntos de datos posee 18 comunidades autónomas, que serán las llamadas observaciones, y 18 variables tras la transformación y depuración de estas en el apartado anterior.

El procedimiento a seguir para construir el indicador global será mediante la creación de subgrupos de variables según la temática para reducir el número de  $p$ .

Se ha dividido el conjunto de variables en dos grupos temáticos: el subgrupo suministro y el subgrupo alcantarillado y depuración. Las razones de esta subdivisión se argumentan gracias al informe del INE sobre *Estadística sobre el Suministro y Saneamiento del Agua – Año 2020*, en el que realiza un estudio sobre la mayoría de las variables escogidas en este trabajo. Tal y como se comenta en el objetivo del informe, este se centra en proporcionar métricas sobre las actividades del ciclo integral del agua, compuesto por el abastecimiento de agua (suministro) y el saneamiento (alcantarillado y depuración de aguas residuales).

A continuación, en la *tabla 3* se muestran las variables que conforman cada subgrupo. El primer grupo temático de suministro está formado por 13 variables y el grupo de alcantarillado y depuración por 5 variables.

**Tabla 3**

*Subgrupos y variables que pertenecen a cada subgrupo.*

<b>Subgrupo suministro</b>	X3: Porcentaje de población con suministro con cobertura del Sistema Nacional de Aguas de Consumo
	X5: Coste Suministro
	X7: Importe facturado por el agua suministrada por habitante
	X9: Longitud red de suministro por habitante
	X11: Pérdidas reales por km de suministro
	X12: Importe total de la inversión en servicios de suministro por habitante
	X13: Número de tanques de tormenta
	X14: Volumen total de agua registrada y distribuida en sectores económicos por habitante
	X15: Volumen total de agua registrada y distribuida en hogares por habitante
	X16: Volumen total de agua registrada y distribuida en consumos municipales por habitante

	X17: Porcentaje pérdidas reales sobre el volumen de agua suministrada X18: Volumen de agua disponible no potabilizada por habitante X19: Volumen de agua disponible para su potabilización por habitante
<b>Subgrupo alcantarillado y depuración</b>	X6: Coste Alcantarillado y depuración X8: Importe facturado por alcantarillado y depuración por habitante X10: Longitud red de alcantarillado por habitante X20: Volumen de aguas residuales tratadas por habitante X22: Volumen de lodos generados en el tratamiento de aguas residuales por habitante

Al disponer de 3 períodos temporales o poblaciones diferentes (2016, 2018, 2020) y 2 grupos temáticos distintos (Subgrupo Suministro, Subgrupo Alcantarillado y depuración), se disponen de 3 conjuntos de individuos sobre los que se han medido las variables correspondientes al grupo Suministro y otros 3 conjuntos de observaciones pertenecientes a las variables del grupo Alcantarillado y depuración.

La condición necesaria para poder obtener el indicador del grupo Suministro y del grupo Alcantarillado y depuración a partir de PCA, es que las matrices de covarianzas poblacionales de cada período sean iguales. Esto se contrasta mediante el test de Barlett.

Se parte de tres matrices de datos  $X_{2016}$ ,  $X_{2018}$ ,  $X_{2020}$  de tamaño  $n_{\alpha} \times p$  siendo  $\alpha = 2016, 2018, 2020$ , de las que se supone que provienen de un vector aleatorio con ley Normal multivariante  $N_p(\mu_{\alpha}, \Sigma_{\alpha})$ . A continuación, se calcula la *pooled within groups covariance matrix*, que se define como la media ponderada de las matrices de covarianzas particulares de cada período, siendo:  $S = \frac{1}{n} \sum_{\alpha=2016}^{2020} n_{\alpha} S_{\alpha}$ , donde  $n = n_{2016} + n_{2018} + n_{2020}$

Por tanto, se lleva a cabo el test de Barlett, que tiene como hipótesis nula que las covarianzas poblacionales sean iguales:  $H_0 = \Sigma_{2016} = \Sigma_{2018} = \Sigma_{2020}$ . El estadístico por contrastar es el basado en la razón de verosimilitudes: 
$$\lambda_R = \frac{|S_{2016}|^{\frac{n_{2016}}{2}} + |S_{2018}|^{\frac{n_{2018}}{2}} + |S_{2020}|^{\frac{n_{2020}}{2}}}{|S|^{\frac{n}{2}}}$$

Si el estadístico siguiente es significativo,  $H_0$  se rechaza:  $-2 \log(\lambda_R) = n \log(|S|) - \sum_{\alpha=2016}^{2020} n_{\alpha} \log(|S_{\alpha}|) \sim \chi_q^2$ , sigue una distribución chi-cuadrado donde los grados de libertad son  $q = \frac{(3-1)p(p+1)}{2}$

Es fundamental que se acepte la hipótesis nula para poderse aplicar PCA a partir de la matriz de covarianzas común (S) o la matriz de correlaciones común, además de que, de esta manera, se pueden determinar unos ejes de representación comunes para los 3 períodos del estudio.

La matriz de correlaciones común o *pooled within groups correlation matrix*, se denomina como  $R = D_s^{-1} S D_s^{-1}$ , siendo  $D_s = \text{diag}(s_1, s_2, \dots, s_p)$  la matriz diagonal con las desviaciones típicas de las  $p$  variables. (Baíllo, A. and Grané, 2007).

Por tanto, tras la aplicación del test de Barlett para el grupo Suministro, se obtiene que el valor de la chi-cuadrado es 176,84 con 182 grados de libertad, y un p-valor de 0,59, siendo mayor que un nivel de significación de  $\alpha = 0,05$ . De tal forma que se acepta  $H_0$  y se puede calcular el indicador mediante PCA.

Asimismo, el test de Barlett para el grupo Alcantarillado y depuración muestra un valor de la chi-cuadrado de 9,52 con  $q = 30$  y un p-valor de 0,99987, motivos por los cuales se vuelve a aceptar  $H_0$  y se puede aplicar PCA en este grupo también.

Gracias a que los resultados anteriormente obtenidos en el test de Barlett para el grupo Suministro y el grupo Alcantarillado y depuración concluyen que las matrices de covarianzas son iguales, es posible aplicar el análisis de componentes principales para cada grupo temático teniendo en cuenta los 3 períodos temporales.

En una primera fase, se empleará el PCA para construir un subindicador correspondiente al grupo Suministro mediante la matriz de correlaciones común (R). Posteriormente, se repetirá el proceso con el otro grupo de Alcantarillado y depuración, obteniendo 3 valores del subindicador del grupo por CCAA, un valor por cada año (2016, 2018, 2020). Se empleará la matriz de correlaciones (y no la de covarianzas) común en ambos casos de grupos temáticos de variables, debido a que las varianzas y las unidades de medida entre las variables de cada grupo difieren entre sí.

#### **4.1.1. PCA: grupo Suministro**

En la *tabla A.2.* del apéndice se aprecian cómo las varianzas de las 13 variables que componen el grupo Suministro son distintas entre sí. Es más, resulta llamativo que la variable X17 tenga una varianza significativamente superior al resto. Esto es porque es la única variable del grupo a la que no se le ha aplicado una transformación logarítmica que homogenice su varianza, por lo que, si se calculase el PCA con la matriz de covarianzas común, la primera componente principal explicaría gran variabilidad del sistema y sería esencialmente X17. Por tanto, se decide que se empleará la matriz de correlaciones común (R) para este primer análisis de componentes principales del grupo Suministro.

En la siguiente *tabla 4*, se muestra la varianza explicada por cada componente principal, que son los autovalores, y la variabilidad explicada acumulada por cada uno. La suma total de las cifras de los autovalores es igual al número de variables iniciales del conjunto de datos, por lo que un autovalor superior a 1 significará que esa componente principal explica más varianza de la que explicaría una variable por sí misma.



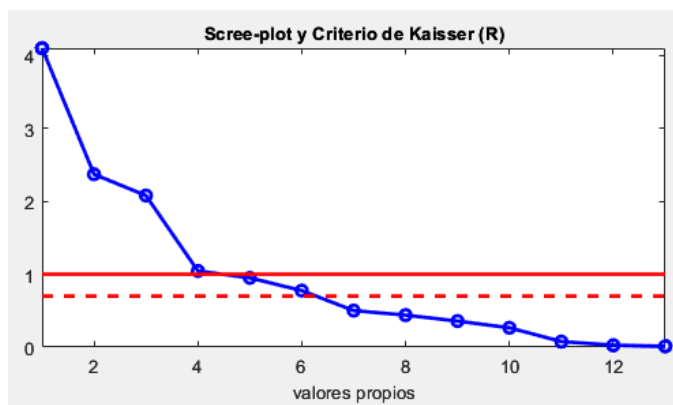
**Tabla 4**

*Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del grupo Suministro*

Componentes principales	Autovalores	Variabilidad explicada acumulada
PC1	4,10	31,55%
PC2	2,37	49,79%
PC3	2,08	65,78%
PC4	1,04	73,82%
PC5	0,95	81,13%
PC6	0,78	87,09%
PC7	0,50	90,95%
PC8	0,44	94,33%
PC9	0,36	97,07%
PC10	0,27	99,11%
PC11	0,08	99,70%
PC12	0,03	99,91%
PC13	0,01	100,00%

**Figura 6**

*Autovalores del PCA del grupo Suministro*



Los resultados que se obtienen sugieren escoger 6 componentes principales, que explican el 87,09% de la variabilidad total. Es decir, reducimos de 13 variables a 6 componentes principales, que resumen la información original en un porcentaje bastante alto. Esta decisión se argumenta a partir del criterio de Jolliffe, que se queda con los autovalores mayores que 0.7, se plasma como la línea horizontal roja discontinua en la *figura 6*. En este caso, si se hubiese aplicado el criterio de Kaiser, que opta por aquellos autovalores superiores a 1 (este criterio lo

marca la línea horizontal roja continua de la *figura 6*), se hubiesen seleccionado 4 componentes; en cambio, estos explicarían únicamente el 73,82% de la variabilidad de los datos. Es decir, se eligen 2 componentes más a cambio de explicar más de un 13% de la variabilidad total. Atendiendo a los valores de la variabilidad acumulada de la tabla, se aprecia claramente como a partir de la PC6 los valores se aplanan; lo que contradice en cierta manera con lo que se observa en el gráfico, ya que el punto del “*elbow*” se sitúa en PC4. Teniendo en cuenta todos los criterios, se escogerán 6 componentes en base al criterio de Jolliffe.

A partir de los autovalores (se encuentran en la *tabla A.3.* del anexo), las ecuaciones de las 6 primeras componentes principales del grupo Suministro son las siguientes:

$$PC1 = 0,28 X3 - 0,28 X5 - 0,26 X7 - 0,06 X9 + 0,39 X11 - 0,20 X12 - 0,29 X13 \\ + 0,14 X14 - 0,02 X15 + 0,16 X16 + 0,35 X17 + 0,39 X18 + 0,41 X19$$

$$PC2 = -0,02 X3 + 0,45 X5 + 0,49 X7 + 0,17 X9 + 0,22 X11 + 0,12 X12 - 0,35 X13 \\ - 0,22 X14 + 0,29 X15 - 0,15 X16 + 0,34 X17 - 0,13 X18 + 0,24 X19$$

$$PC3 = -0,06 X3 - 0,20 X5 - 0,11 X7 + 0,50 X9 - 0,20 X11 - 0,42 X12 + 0,14 X13 \\ + 0,27 X14 + 0,44 X15 - 0,37 X16 + 0,07 X17 - 0,14 X18 + 0,16 X19$$

$$PC4 = 0,51 X3 - 0,01 X5 - 0,08 X7 + 0,33 X9 - 0,18 X11 - 0,10 X12 + 0,23 X13 \\ - 0,64 X14 + 0,11 X15 + 0,30 X16 + 0,09 X17 + 0,06 X18 - 0,09 X19$$

$$PC5 = -0,04 X3 + 0,09 X5 - 0,07 X7 - 0,24 X9 + 0,24 X11 - 0,46 X12 + 0,21 X13 \\ - 0,30 X14 - 0,36 X15 - 0,51 X16 + 0,31 X17 - 0,09 X18 - 0,19 X19$$

$$PC6 = -0,63 X3 - 0,11 X5 - 0,25 X7 + 0,44 X9 + 0,01 X11 + 0,03 X12 - 0,19 X13 \\ - 0,21 X14 - 0,30 X15 + 0,28 X16 + 0,26 X17 - 0,12 X18 - 0,04 X19$$

Para interpretar cada componente principal y conocer qué variables influyen más en cada uno de ellos, hay que fijarse en el coeficiente que acompaña a cada variable (que es el autovector). Cuanto más grande sea en valor absoluto, más peso tendrá sobre el componente principal. Su signo indicará si contribuye de forma positiva o negativa, interpretando así qué características explica cada componente principal.

En el caso del grupo Suministro:

- La primera componente, explica un 31,55% de la variabilidad total como se puede observar en la *tabla 4*. Destacan los pesos positivos en torno a 0,3 y 0,4 de las variables X3 (Porcentaje de población con suministro del Sistema Nacional de Aguas de Consumo), X11 (Pérdidas reales por km en la red de suministro), X17 (Porcentaje de pérdidas reales sobre el volumen de agua suministrada); X18 y X19, que son los volúmenes de agua disponible no potable y potable por habitante. Por tanto, se puede interpretar esta primera componente principal como un indicador de cantidad de agua disponible por habitante, que mide lo eficaz que es la CCAA a la hora de suministrar agua gracias a X3. Sin embargo, al tratar con grandes volúmenes de agua, al

mismo tiempo obtiene grandes pérdidas (aumenta PC1 al aumentar X11 y X17), por lo que se puede intuir que no utiliza los recursos de forma eficiente. En resumen, refleja un suministro eficaz, pero no eficiente.

- Añadiendo una segunda componente, ambas explican un 49,79% de la variabilidad total. Prevalen frente al resto la variable X5 (Coste unitario del suministro) ponderada negativamente, y X7 (Importe facturado por habitante por el agua suministrada), ambas con un peso considerablemente positivo.

De modo que, se puede deducir que PC2 mide el coste unitario del suministro, que tiene una relación directamente proporcional con el importe facturado, pero actúa como consecuencia del coste (a mayor coste del suministro, mayor es el importe facturado). Debido a que el objetivo siempre es minimizar costes, cuanto más negativo sea el valor de esta componente, implicará que su coste es menor (como consecuencia, el importe facturado por habitante por el suministro también será menor). <sup>1</sup>

<sup>1</sup> NOTA: Para facilitar la interpretación, se han cambiado los signos de todos los autovectores de la segunda componente principal, multiplicándolos por -1. De forma que gráficamente, los valores más hacia arriba en el eje vertical en la figura 7 sean los más positivos, facilitando la interpretación para que las mejores CCAA respecto a esta segunda componente se sitúen lo más arriba posible. Manteniendo así también la coherencia en la formación del subindicador del grupo suministro, ya que las puntuaciones de las mejores CCAA's en la PC2 contribuirán de forma positiva.

$$^1 PC2 = 0,02 X3 - 0,45 X5 - 0,49 X7 - 0,17 X9 - 0,22 X11 - 0,12 X12 + 0,35 X13 + 0,22 X14 - 0,29 X15 + 0,15 X16 - 0,34 X17 + 0,13 X18 - 0,24 X19$$

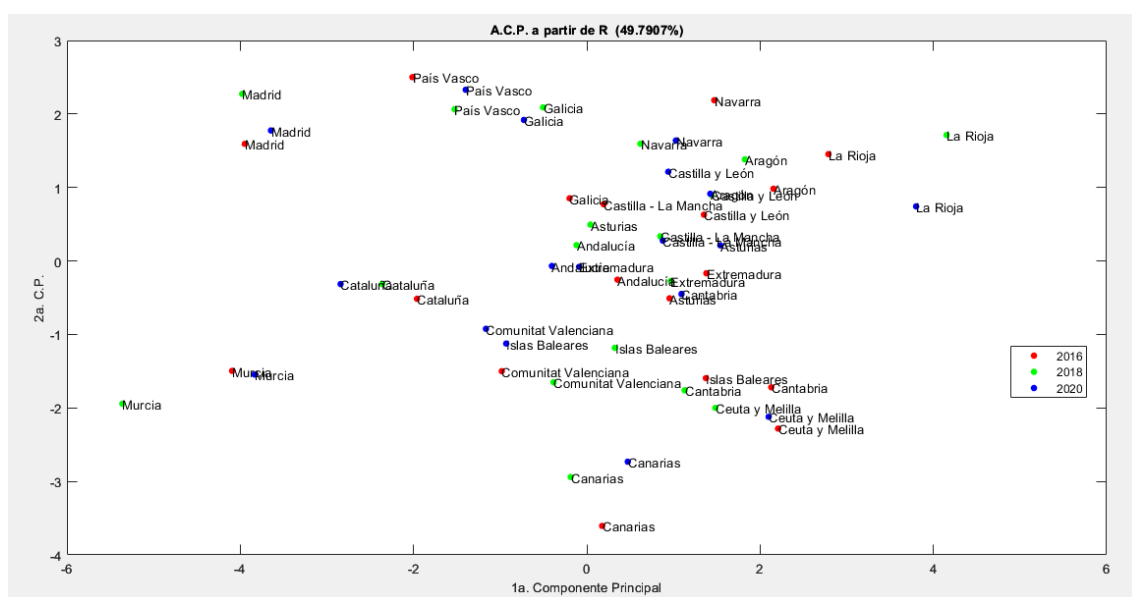
- Por su parte, la tercera componente principal ofrece pesos positivos en las variables X9 (Longitud red de suministro por habitante) y X15 (Volumen total de agua distribuida en hogares), por lo que se puede interpretar como un indicador de las redes de suministro con más km que conectan más hogares.
- Teniendo en cuenta hasta la cuarta componente, se explicaría un 73,82%. En ella sobresale la influencia negativa de la variable X14 (Volumen total de agua distribuida en sectores económicos) y la positiva de X3. De forma que los valores máximos de esta componente indicarán las CCAA's que menos agua reparten a empresas y más población posee suministrada.
- La quinta componente refleja aquellas comunidades que menos invierten en servicios de suministro y, por tanto, menos volumen de agua distribuyen por usuario.
- Por último, con la sexta componente se explica un 87,09% y en concreto esta componente manifiesta las comunidades con más longitud de red de suministro y menos población suministrada.

A continuación, en la *figura 7*, se muestra la interpretación gráfica de las dos primeras componentes sobre un plano bidimensional. Siendo el eje horizontal la puntuación de un individuo en la primera componente, y el eje vertical la puntuación de la segunda componente para una determinada CCAA en un período de tiempo específico.

En el apéndice, se proporciona toda la información útil sobre los componentes principales, tal como los valores de los autovectores o las puntuaciones de cada CCAA en cada año sobre las componentes principales obtenidas en el PCA del grupo Suministro.

**Figura 7**

*Interpretación gráfica de las dos primeras componentes del grupo Suministro*



Esto permite analizar la situación de cada comunidad autónoma respecto al grupo Suministro y manifiesta lo que se venía interpretando de las dos primeras componentes. Como se señalaba, cuanto mayor era el valor de la primera componente mejor posición respecto a un suministro eficaz, pero no eficiente. Se deduce por tanto que la mejor CCAA en estos aspectos es La Rioja, mientras que a la cola observamos a Cataluña, Comunidad de Madrid y Murcia. Es decir, teniendo en cuenta las siguientes variables de forma global (PC1), La Rioja es la CCAA con mayor porcentaje de población suministrada, mayor volumen por habitante de agua potable y no potable y mayores pérdidas por km de red de suministro y por el agua suministrada.

Con respecto a la segunda componente principal, los valores positivos ahora indican un menor coste (y como consecuencia un menor importe facturado por habitante por el agua suministrada). Por tanto, la mejor posición la tendrán los valores más positivos de esta PC2, que en este caso corresponden a País Vasco, Comunidad de Madrid y Navarra, ya que tendrán un menor coste, lo cual es el objetivo. De la misma manera, Canarias se sitúa como la peor situada en este aspecto, con un coste unitario de suministro muy elevado.

En definitiva, cuanto más cerca de la esquina superior derecha, mejor puntuación obtendrá en el subindicador del grupo Suministro.

Es llamativo como no hay grandes diferencias entre los 3 periodos temporales (años 2016, 2018 y 2020), lo que va en consonancia con los resultados obtenidos anteriormente en el test de Barlett, de que las matrices de covarianzas poblacionales de cada periodo son significativamente iguales.

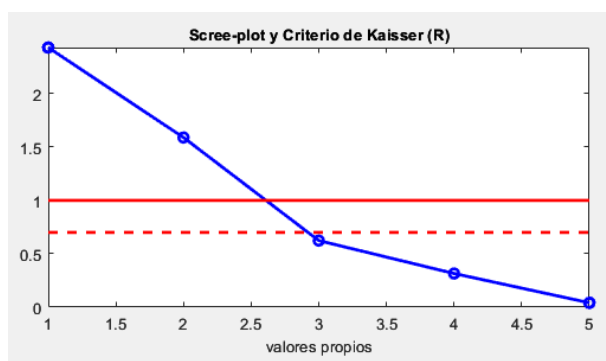
#### 4.1.2. PCA: grupo Alcantarillado y depuración

El grupo temático Alcantarillado y depuración está compuesto por 5 variables. Se optará de nuevo por utilizar la matriz de correlaciones común ( $R$ ) debido a que las unidades de medida de las variables son distintas entre sí, lo que se demuestra al observar las varianzas propias de cada variable en la *tabla A.5.* del apéndice. La varianza de X10 comparada al resto, lo que implicaría que al utilizar la matriz de covarianzas común ( $S$ ), esta variable que indica la longitud de la red de alcantarillado por habitante, tenga demasiada relevancia en el estudio del PCA.

Siguiendo el criterio de Jolliffe de que los autovalores sean superiores a 0,7, y cumpliendo de forma simultánea el criterio de Kaiser (los autovalores sean superiores a la unidad), la figura 8 sugiere tomar 2 componentes principales, que explican un 80,41% de la variabilidad total del grupo Alcantarillado y depuración, como se denota en la *tabla 5.*

**Figura 8**

*Autovalores del PCA del grupo Alcantarillado y depuración*



**Tabla 5**

*Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del grupo Alcantarillado y depuración*

Componentes principales	Autovalores	Variabilidad explicada acumulada
PC1	2,43	48,63%
PC2	1,59	80,41%
PC3	0,62	92,87%
PC4	0,32	99,18%
PC5	0,04	100,00%

Desde sus correspondientes autovalores (encontrados en la *tabla A.6.* del apéndice), se pueden calcular las ecuaciones de las dos primeras componentes principales del grupo Alcantarillado y depuración, que son las siguientes:

$$PC1 = -0,49 X6 - 0,58 X8 - 0,18 X10 - 0,34 X20 - 0,53 X22$$

$$PC2 = -0,50 X6 - 0,28 X8 + 0,63 X10 + 0,46 X20 + 0,25 X22$$

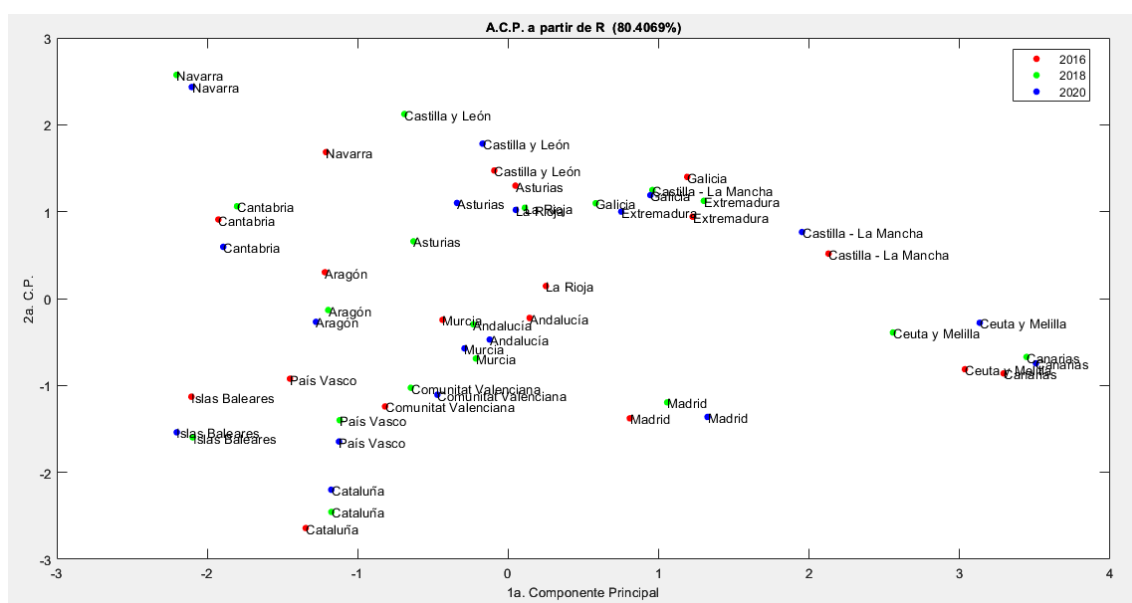
Interpretando los pesos de las variables en cada componente:

- La primera componente principal explica el 48,63% de la variabilidad total, lo que resulta casi la mitad del grupo Alcantarillado y depuración. Todas las variables contribuyen de forma negativa a esta componente, especialmente con un destacado peso las variables X6 (Coste unitario de alcantarillado y depuración), X8 (Importe facturado por habitante por el alcantarillado y depuración) y X22 (Volumen de lodos generados por habitante en el tratamiento de aguas residuales). Por tanto, cuanto mayor sea el valor de estas variables en cada CCAA, más hacia la izquierda del eje se situará la primera componente. Al ser el objetivo minimizar todas estas variables, estar lo más a la derecha del eje indicará una mejor situación. Se puede definir como un indicador de bonanza del grupo Alcantarillado y depuración, ya que la que mayor valoración obtenga de esta primera componente será la CCAA de forma conjunta con menor coste unitario de alcantarillado (y menor importe facturado por el alcantarillado, es una consecuencia del coste) y menor volumen de lodos generados por habitante.
- Junto a la segunda componente, es capaz de explicar de forma conjunta el 80,41% de la variabilidad de las cinco variables iniciales. En esta componente, está muy presente la variable X10 que indica la longitud de la red de alcantarillado por habitante, aportando un peso positivo a la PC2. Así mismo, influye de manera negativa el coste del alcantarillado.

El objetivo de cada CCAA debe ser incrementar la longitud de red de alcantarillado por habitante al menor coste posible, de forma que, cuanto mayor sea el valor de esta eficiencia de la longitud de red de alcantarillado, mejor puntuación obtendrá en el eje vertical cada CCAA.

**Figura 9**

*Interpretación gráfica de las dos primeras componentes del grupo Alcantarillado y depuración*



En la *figura 9* se puede detectar que las CCAA's con mejor posición tienen con respecto a la primera componente (son las que están más a la derecha en el eje X) se tratan de Canarias, Ceuta y Melilla y Castilla – La Mancha. Son aquellas con menor coste unitario de alcantarillado, menor volumen de aguas residuales por habitante y menor volumen de lodos generados por habitante.

La segunda componente principal, como se ha indicado anteriormente, refleja las CCAA's con mayor eficiencia de la longitud de red de alcantarillado lo más arriba posible del eje vertical. Es decir, poseen más km de red de alcantarillado al menor coste unitario. Estas son Navarra y Castilla y León. Al final de la lista, encontramos a Cataluña, Islas Baleares, País Vasco y Madrid.

En conclusión, podemos interpretar que el eje horizontal está ordenado de menor a mayor por un índice de bondad respecto a este grupo temático; y el eje vertical de menor a mayor índice de eficiencia de la longitud de red de alcantarillado.

Como resumen de la primera fase, se han reducido las dimensiones del grupo Suministro con 6 componentes principales que explican el 87,09% de la variabilidad total. Por otra parte, del grupo Alcantarillado se han escogido 2 componentes principales que describen el 80,41% de la variabilidad total.

Para cada grupo, se ha calculado un subindicador teniendo en cuenta las varianzas explicadas por cada componente de forma ponderada. Ambos subindicadores muestran una calificación de cada CCAA en cada grupo correspondiente, que se analiza con más claridad con el indicador estandarizado.

Los top 3 mayores valores en el grupo Suministro son: La Rioja (en el año 2018) con un 100, La Rioja (en el año 2020) con un valor de 92,56 y Aragón (en el año 2016) con una valoración de 87,58 . Y las que peor puntuación obtienen son Región de Murcia, con los tres valores más bajos, 0 en el año 2018, 9,83 en el 2016 y 27,42 en el año 2020; seguido de comunidades como Comunidad de Madrid y Cataluña que están a la cola de este ranking.

Respecto al grupo Alcantarillado y depuración, las top 3 mejores valoraciones son: Canarias (en 2020) con una puntuación del indicador estandarizado de 100 , seguido por Canarias (en el año 2018) con 99,78 y, en el último escalón, se encuentra Ceuta y Melilla (en 2020) con 98,90 . Como calificaciones más bajas del ranking de este grupo, se sitúan Islas Baleares (en 2020) con la puntuación más baja 0, Islas Baleares (en 2018) obteniendo 1,09 puntos, y terminando, Cataluña en 2016 con una puntuación de 2,19.

## 4.2. Construcción del Indicador Global

Con esto concluye la primera fase, y comienza una segunda fase, en la que se formará un nuevo conjunto de datos compuesto por los subindicadores de cada grupo temático y todos los períodos temporales. Es decir, este dispondrá de 54 observaciones (18 CCAA's por cada año 2016, 2018 y 2020), y 2 variables, que son los subindicadores sin estandarizar que se acaban de crear de cada grupo temático (Grupo Suministro y Grupo Alcantarillado y depuración).

Para comprobar si las matrices de covarianzas poblacionales de cada período temporal son iguales sobre este nuevo conjunto de datos más reducido, es necesaria la aplicación del test de Barlett. Los resultados son una chi-cuadrado igual a 0,9036 con 6 grados de libertad y un p-valor de 0,989. Por tanto, concluyen que se acepta  $H_0$  y se puede calcular el indicador global mediante PCA, a través de la matriz de covarianzas o de correlaciones común.

Se opta por utilizar la matriz de correlaciones común  $R = \begin{bmatrix} 1 & 0,19 \\ 0,19 & 1 \end{bmatrix}$ , que se calcula de la forma:  $R = D_s^{-1} S D_s^{-1}$  , siendo  $D_s = \text{diag}(s_1, s_2, \dots, s_p)$  la matriz diagonal con las desviaciones típicas de las  $p$  variables. La razón es que las varianzas de nuevas variables creadas son diferentes entre sí, por lo que conviene utilizar la  $R$ , ya que estandariza los datos. De forma que, ahora, la PC1 de  $R$ , proporciona el mismo peso en valor absoluto a cada una de las variables Suministro y Alcantarillado y depuración.

Siendo  $e_1 = (0,71; 0,71)'$  los autovectores de PC1 y  $\lambda_1 = 1,19$  el autovalor de PC1; y  $Z_1$  y  $Z_2$  las variables Suministro y Alcantarillado y depuración estandarizadas, respectivamente:



$$\text{corr}(PC1, Z_1) = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{r_{11}}} = \frac{0,71 \cdot \sqrt{1,19}}{\sqrt{1}} = 0,775 \quad \text{corr}(PC1, Z_2) = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{r_{22}}} = \frac{0,71 \cdot \sqrt{1,19}}{\sqrt{1}} = 0,775$$

(Baíllo, A. and Grané, 2007).

De la matriz de correlaciones común se puede concluir que las variables Suministro y Alcantarillado y depuración están prácticamente incorreladas entre sí, su correlación es de 0,19.

Además, se observa una alta correlación positiva entre la variable Suministro y la primera componente, así como entre Alcantarillado y depuración y PC1, de 0,775. Es más, los valores obtenidos a partir de la primera componente principal coincidirán con los del indicador global. Por tanto, otra forma de interpretarlo, es que se acaba de calcular la correlación entre el indicador global y las variables compuestas por los subindicadores temáticos.

Como la primera componente principal explica un 59,66% de la variabilidad (según se observa en la *tabla 6*) será el valor del indicador global. Habrá 3 valores del indicador global por CCAA, un valor por cada año (2016, 2018, 2020).

**Tabla 6**

*Autovalores y porcentaje de variabilidad explicada acumulada de cada componente principal del nuevo conjunto de datos.*

Componentes principales	Autovalores	Variabilidad explicada acumulada
PC1	1,19	59,66%
PC2	0,81	100,00%

**Tabla 7**

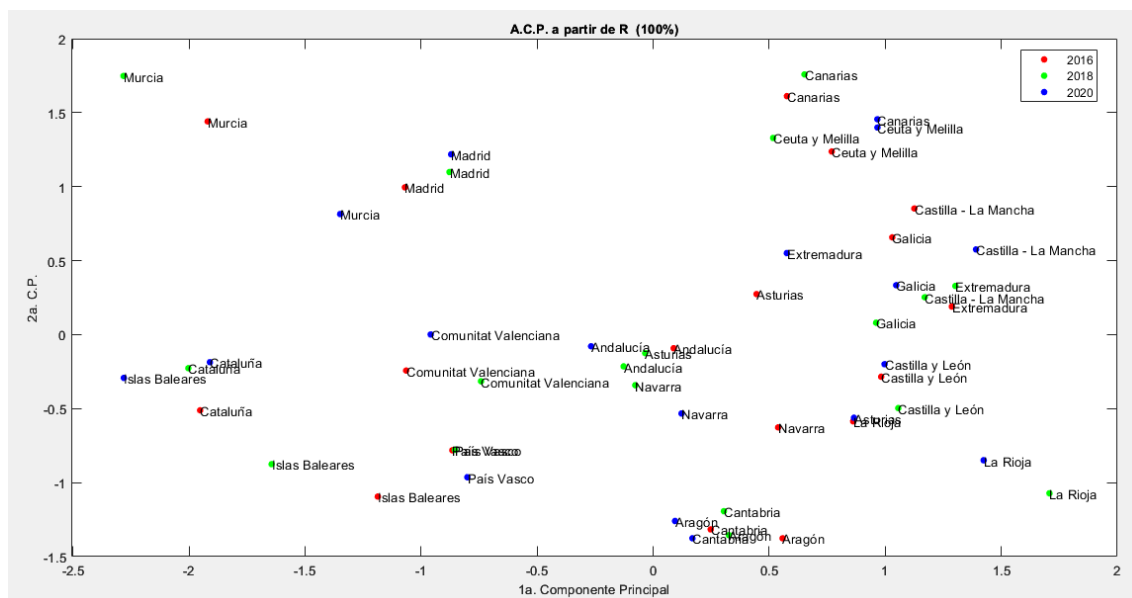
*Autovectores de cada componente principal del nuevo conjunto de datos reducido*

	PC1	PC2
Suministro	0,71	- 0,71
Alcantarillado y depuración	0,71	0,71

$$PC1 = 0,71 \text{ Suministro} + 0,71 \text{ Alcantarillado y depuración}$$

**Figura 10**

*Interpretación gráfica de las dos primeras componentes del nuevo conjunto de datos reducido*



Haciendo énfasis en el eje horizontal, coincidente con las puntuaciones del PC1 (explica un 59,66% de la variabilidad total) y con los valores del indicador global, se puede establecer un ranking de CCAA y agrupar las mismas en escalones.

En el primer escalón, como las CCAA que más se encuentran a la derecha del eje X, se encuentra La Rioja (únicamente en los años 2018 y 2020) a la cabeza de la lista, seguida de Castilla-La Mancha, Extremadura (solo en los años 2016 y 2018), Galicia y Castilla y León.

En el último escalón, lo más a la izquierda del eje X, se sitúan desmarcadas del resto Región de Murcia, Islas Baleares y Cataluña.

Por último, en una zona media, se hallan diferenciadas de los extremos el resto de las comunidades autónomas. Dentro de la que se encuentran Comunidad de Madrid, Comunitat Valenciana, País Vasco, Andalucía, Navarra, Aragón, Cantabria, Asturias, Canarias y Ceuta y Melilla, en orden ascendente.

**Tabla 8**

*Puntuaciones del Indicador global por CCAA y por año*

CCAA	AÑO	PC1/INDICADOR GLOBAL	INDICADOR ESTANDARIZADO	GLOBAL
01 Andalucía	2016	0,09	59,42	
02 Aragón	2016	0,56	71,18	
03 Asturias, Principado de	2016	0,45	68,37	
04 Balears, Illes	2016	-1,19	27,46	
05 Canarias	2016	0,58	71,65	
06 Cantabria	2016	0,25	63,42	
07 Castilla y León	2016	0,98	81,83	
08 Castilla - La Mancha	2016	1,13	85,41	
09 Cataluña	2016	-1,95	8,28	
10 Comunitat Valenciana	2016	-1,06	30,53	
11 Extremadura	2016	1,29	89,46	
12 Galicia	2016	1,03	83,02	
13 Madrid, Comunidad de	2016	-1,07	30,40	
14 Murcia, Región de	2016	-1,92	9,09	
15 Navarra, Comunidad Foral de	2016	0,54	70,70	
16 País Vasco	2016	-0,87	35,51	
17 Rioja, La	2016	0,86	78,81	
18 Ceuta y Melilla	2016	0,77	76,49	
01 Andalucía	2018	-0,13	54,01	
02 Aragón	2018	0,33	65,38	
03 Asturias, Principado de	2018	-0,03	56,34	
04 Balears, Illes	2018	-1,64	16,00	
05 Canarias	2018	0,65	73,54	
06 Cantabria	2018	0,30	64,84	
07 Castilla y León	2018	1,06	83,67	
08 Castilla - La Mancha	2018	1,17	86,52	
09 Cataluña	2018	-2,00	7,01	
10 Comunitat Valenciana	2018	-0,74	38,61	
11 Extremadura	2018	1,30	89,84	
12 Galicia	2018	0,96	81,28	
13 Madrid, Comunidad de	2018	-0,88	35,21	
14 Murcia, Región de	2018	-2,28	0,00	
15 Navarra, Comunidad Foral de	2018	-0,08	55,27	
16 País Vasco	2018	-0,85	35,80	

<b>17 Rioja, La</b>	2018	1,71	100,00
<b>18 Ceuta y Melilla</b>	2018	0,52	70,16
<b>01 Andalucía</b>	2020	-0,27	50,50
<b>02 Aragón</b>	2020	0,09	59,57
<b>03 Asturias, Principado de</b>	2020	0,87	78,89
<b>04 Balears, Illes</b>	2020	-2,28	0,03
<b>05 Canarias</b>	2020	0,97	81,41
<b>06 Cantabria</b>	2020	0,17	61,44
<b>07 Castilla y León</b>	2020	1,00	82,19
<b>08 Castilla - La Mancha</b>	2020	1,39	92,07
<b>09 Cataluña</b>	2020	-1,91	9,33
<b>10 Comunitat Valenciana</b>	2020	-0,96	33,18
<b>11 Extremadura</b>	2020	0,58	71,64
<b>12 Galicia</b>	2020	1,05	83,45
<b>13 Madrid, Comunidad de</b>	2020	-0,87	35,39
<b>14 Murcia, Región de</b>	2020	-1,35	23,40
<b>15 Navarra, Comunidad Foral de</b>	2020	0,12	60,26
<b>16 País Vasco</b>	2020	-0,80	37,14
<b>17 Rioja, La</b>	2020	1,42	92,89
<b>18 Ceuta y Melilla</b>	2020	0,97	81,43

En comparación con las visualizaciones en los grupos temáticos, se aprecia una ligera diferencia entre años (sigue cumpliendo el test de Barlett de que las matrices de covarianzas poblacionales son iguales, como se ha visto anteriormente) que se analizará a continuación a través de la representación de los valores del indicador global mediante mapas geográficos y diagramas de caja.

## 5. ANÁLISIS DE LOS RESULTADOS

### 5.1. Visualización del Indicador Global

**Figura 11**

*Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2016*



**Figura 12**

*Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2018*



**Figura 13**

*Mapa geográfico de las CCAA españolas con los valores (0-100) del Indicador global en 2020.*



Atendiendo al degradado de color azul que se observa en los mapas geográficos, se pueden estudiar las variaciones más significativas de un año a otro.

Por ejemplo, del año 2016 a 2018, los crecimientos más notables se observan La Rioja, ya que pasa de un valor de 78,81 a 100, lo que implica un aumento del 26,89% de su valor. También destacan la Comunitat Valenciana, que crece un 26,49% (sube de 30,53 a 38,61) y la Comunidad de Madrid, con una subida del 15,82% (de 30,40 en 2016 a 35,21 en 2018). Como bajadas importantes, resaltan Murcia (que baja de 9,09 en 2016 a 0 en 2018) o Islas Baleares con una caída del 41,75% (pasa de 27,46 a 16,00).

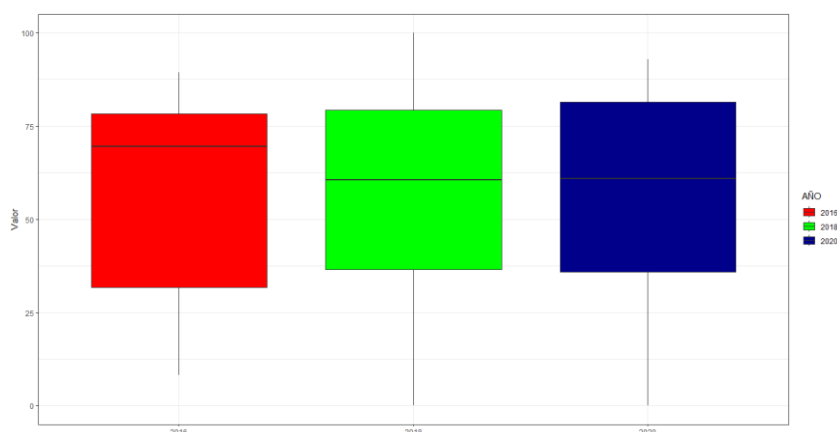
Navarra y Asturias son dos comunidades que siguen un compartimiento de bajada de 2016 a 2018 para posteriormente crecer de 2018 a 2020. Por un lado, Navarra desciende un 21,83% en el primer tramo (de 70,70 en 2016 a 55,27 en 2018) para crecer un 9,04% en el segundo tramo (de 55,27 en 2018 a 60,26 en 2020). Por su parte, Asturias decrece un 17,60% de 2016 a 2018 (pasa de 68,37 a 56,34) para alcanzar su punto más álgido en 2020 con una subida de un 40,04% y un valor del indicador de 78,89. Comparten este mismo patrón de comportamiento, las siguientes comunidades: Cataluña, Ceuta y Melilla, Galicia y Región de Murcia. Es decir, todas estas comunidades tienen en común que su valoración más baja se obtuvo en 2018.

Con respecto al segundo tramo, de 2018 a 2020, crece notablemente Canarias un 10,70%, continuando con su tendencia alcista a lo largo de todos los años. Otras CCAA's que poseen un crecimiento periódico prolongado son Castilla-La Mancha, Comunidad de Madrid y País Vasco.

Por el contrario, como bajadas significativas de 2018 a 2020 se encuentran Islas Baleares siguiendo con su descenso asiduo (pasa de 16,00 a 0,03) y Extremadura que sufre una caída brusca del 20,26% (de 89,84 en 2018 a 71,64 en 2020). Cabe mencionar las CCAA's con un decrecimiento continuado a lo largo de los períodos temporales, que además de Islas Baleares son: Andalucía y Aragón.

#### Figura 14

*Comparativa boxplots del Indicador global entre los años 2016, 2018 y 2020*



En las anteriores *figuras 11, 12, 13 y 14*, se puede observar cómo no hay una evolución de mejora clara entre los tres años del espacio temporal llevado a estudio, de hecho, el valor máximo del indicador se registra en 2018 en La Rioja.

Los picos máximos y mínimos del indicador global se encuentran en el año 2018; en contraste con 2016, donde se hallan los valores más concentrados en torno a un valor medio.

Por otro lado, la media alcanza su máximo nivel en 2016, cuando es de 57,83, para posteriormente descender en 2018 a 56,30 ; y finalmente, volver a tomar una subida, pero sin sobrepasar a la de 2016, con un valor medio del indicador global de 57,46.

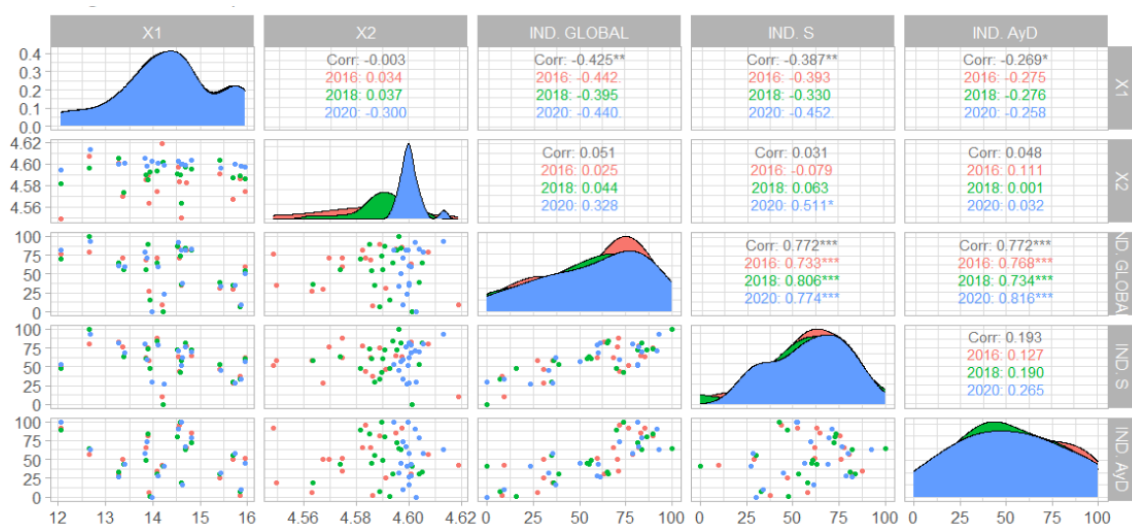
Con respecto a los cuartiles, resulta notable la mediana tan al alza en 2016 ocupando el valor (69,54), denotando que, en este año, los valores del indicador global son generalmente bastante elevados, y sin picos extremos por arriba y por abajo. Además, el valor del tercer cuartil es de 78,23 , lo que implica el 25% de los valores del indicador en 2016 se sitúan en el rango entre 69,54 y 78,23. También, es de considerable interés mencionar que se observa una evolución alcista a lo largo de los años en el Q3.

## 5.2. Relación Indicador Global con Variables Demográficas y Socioeconómicas

En la siguiente *figura 15*, se puede observar una representación de la relación entre las variables demográficas y socioeconómicas X1 (Población) y X2 (IPC del suministro de agua y otros servicios relacionados con la vivienda), respecto al indicador global y también sobre los respectivos subindicadores temáticos de suministro y alcantarillado y depuración.

**Figura 15**

*Diagrama de dispersión matricial con su respectiva matriz de correlaciones, entre las variables X1 y X2 transformadas, el indicador global, y los subindicadores temáticos de suministro y alcantarillado y depuración*



Para empezar, se puede observar como las variables X1 (Población) y X2 (IPC) están prácticamente incorreladas entre sí, su correlación es casi 0, excepto en el año 2020 donde es aproximadamente  $-0,3$ .

Con respecto a la relación del Indicador global con la población, se observa una correlación media de  $-0,425$  entre todos los años por lo que no se trata de una correlación demasiado fuerte. Sin embargo, esta se puede interpretar como que las CCAA's con una mayor población obtienen peores resultados en el indicador global, lo que es coherente con el ODS 6 de un acceso universal al agua y servicios de suministro, ya que ante una mayor población generalmente es lógico que se tenga más dificultades de llegar a todos y cada uno de los habitantes.

La correlación entre el indicador global y la variable del IPC del suministro de agua obtiene un valor de  $0,051$ , lo que indica que ambas están prácticamente incorreladas, no existe correlación lineal entre ellas. De hecho, en el diagrama de puntos se aprecia aleatoriedad.

De la misma manera, la correlación entre los subindicadores temáticos con las variables X1 y X2 sigue la misma interpretación.

El indicador del grupo temático suministro de  $-0,39$  con la variable población y de  $0,031$  con la variable IPC. Este indicador recibía una gran influencia de la variable coste unitario del suministro, por lo que al no tener relación con la variable IPC del Suministro, se puede deducir que no hay ninguna relación de tipo lineal entre el coste y el precio del suministro.

Por su parte, el indicador del grupo temático Alcantarillado y depuración tiene una correlación aún menor con la variable población, de  $-0,27$ ; y tampoco posee relación lineal con el IPC al tener una correlación con esta variable de  $0,048$ .

Por último, cabe destacar que, como se había calculado con anterioridad, existe una alta correlación lineal entre el indicador global y los indicadores temáticos, con un valor de  $0,77$ . Especialmente con el indicador del grupo suministro en 2018, cuando la correlación alcanza un valor de  $0,81$ ; y con el indicador del grupo alcantarillado y depuración en el año 2020, con una correlación de  $0,82$ .



## 6. CONCLUSIONES

A pesar de la importancia del ODS Agua limpia y saneamiento respecto al resto de objetivos de Desarrollo Sostenible de la Agenda 2030, es una de las metas que menos información disponible posee, y la que se encuentra, bien es muy dispersa por toda la información que se puede tratar sobre este ODS, o bien corresponde a países del mundo. En la introducción de este trabajo se ha ofrecido una visión global de la situación mundial sobre este objetivo, pero... ¿cómo es la situación de cada una de las CCAA de España respecto al agua limpia y saneamiento?

El objetivo de este estudio ha sido precisamente crear un indicador global que muestre la situación de cada una de las CCAA españolas respecto a este sexto ODS, en los años 2016, 2018 y 2020.

Para la consecución de ello, se ha necesitado el apoyo de técnicas estadísticas, fundamentalmente el Análisis de Componentes Principales (PCA). La finalidad de esta técnica es reducir la dimensión del conjunto de variables iniciales, formando un número menor de variables llamadas componentes, que maximizan la variabilidad total explicada del conjunto inicial. A partir del PCA, se puede construir un indicador para cada CCAA, que actúa de forma de ranking según las valoraciones que obtiene cada CCAA.

Debido a la numerosidad y heterogéneo significado de las variables de las que se partía, se han formado dos grupos temáticos de variables: grupo Suministro y grupo Alcantarillado y depuración. De forma que se ha construido un subindicador para cada grupo, con el que se ha generado el indicador global.

Por tanto, teniendo en cuenta la interpretación de los componentes principales que conforman cada subindicador, se pueden extraer las siguientes conclusiones.

Con respecto al subindicador creado para el grupo Suministro, los mayores valores son obtenidos por La Rioja, Aragón y Cantabria y son aquellas con mayor eficacia en la distribución de agua, que presentan un menor coste del suministro, al mismo tiempo que poseen las mayores redes de suministro dirigidas a hogares y no a empresas. Por el contrario, la Región de Murcia y la Comunidad de Madrid exhiben las peores calificaciones en suministro debido a que poseen la peor eficacia en la distribución de agua a un mayor coste del suministro.

Por su parte, el subindicador del grupo Alcantarillado y depuración otorga las mejores puntuaciones a Canarias, Ceuta y Melilla y Castilla-La Mancha, ya que son las que obtienen un menor coste de alcantarillado, un volumen menor de aguas residuales y lodos generados por habitante, y son las CCAA que poseen las redes de alcantarillado con mayor longitud. En cambio, las menores valoraciones de este grupo son propiedad de Islas Baleares, Cataluña y País Vasco, serán las que tienen un mayor coste de alcantarillado, un mayor volumen de aguas residuales y lodos generados por habitante y una red de alcantarillado de menor longitud.

Por tanto, una mejora realista de aplicar para las CCAA y que mejoraría su situación respecto a los subindicadores de suministro y alcantarillado y depuración sería una reducción de los costes de suministro y alcantarillado, al mismo tiempo que se aumenten las longitudes de redes de suministro y alcantarillado, para cumplir con el objetivo del suministro universal de agua limpia.

De forma conjunta, el indicador global concede mejores puntuaciones a aquellas CCAA con mejor situación de suministro y de alcantarillado y depuración, cumpliendo con el objetivo principal del trabajo de conseguir un indicador por CCAA en 2016, 2018 y 2020. Por consiguiente, el ranking asigna a La Rioja en 2018 en primer lugar, La Rioja en 2020 en segundo lugar, y Castilla-La Mancha en 2020 en tercer lugar; como las mejores combinaciones de CCAA-año en las que el indicador global del ODS 6 ha tenido una mejor situación.

Ahora bien, como el objetivo es un acceso universal al agua para el año 2030, uno de los grandes objetivos de este estudio era analizar la evolución que viene llevando cada CCAA a lo largo del espacio temporal contemplado (2016, 2018 y 2020). Por este motivo, las comunidades autónomas que presentan un crecimiento progresivo a lo largo de los años son: Canarias, Castilla-La Mancha, Comunidad de Madrid y País Vasco. Sin embargo, aunque esto parezca un indicio de que van por buen camino al estar creciendo, puede ser engañoso. Por ejemplo, en el caso de la Comunidad de Madrid, es cierto que presenta una tendencia al alta, pero sus valores distan bastante de los valores de las CCAA en el top de ranking, por lo que es necesario un impulso, porque de seguir por este camino, a pesar de que es ascendente, no llegaría al objetivo.

Un peor panorama lo tienen aquellas comunidades que presentan un decrecimiento progresivo debido a que sus valores del indicador están cayendo a lo largo de los años. Estas son las siguientes: Islas Baleares, Andalucía y Aragón.

Por último, se estudió si el indicador global tenía relación con variables demográficas y socioeconómicas como la población y el IPC del suministro de agua. Se llegaron a las conclusiones de que hay una correlación negativa débil entre la población y el indicador global, interpretando que una mayor población dificulta el acceso universal

En definitiva, este estudio se ha basado en el análisis de componentes principales, con el objetivo de reducir lo máximo la información hasta construir un indicador final. Sin embargo, sería interesante para futuros trabajos considerar la opción de aplicar métodos de predicción, que simulen cómo acabaría la situación de las diferentes CCAA en el año 2030, y si cumplirían el objetivo del acceso universal al agua limpia para dicha fecha o no. Para la realización de estos métodos, se necesitaría más información para crear una serie de modelos de predicción.

## 7. BIBLIOGRAFÍA

### Código implementado y bases de datos

Baíllo, A. y Grané, A. (2007) 100 problemas de Estadística Multivariante implementados en Matlab. Delta Publicaciones, Madrid.

<https://halweb.uc3m.es/esp/Personal/personas/agrane/libro/100PEM.htm>

INE - Instituto Nacional de Estadística. (s. f.). *Indicadores de la Agenda 2030 para el Desarrollo Sostenible - Objetivo 6. Garantizar la disponibilidad y la gestión sostenible del agua y el saneamiento para todos*. INE. <https://www.ine.es/dyngs/ODS/es/objetivo.htm?id=5003>

### Información académica

*Conferencia de la ONU sobre el agua 2023 | Department of Economic and Social Affairs*. (s. f.). <https://sdgs.un.org/es/conferences/water2023>

*Goal of the Month Editorial | March 2023 - United Nations Sustainable Development*. (2023, March 30). the United Nations. Retrieved July 11, 2023, from <https://www.un.org/sustainabledevelopment/goal-of-the-month-editorial-march-2023/>

Grané Chávez, A., & Albarrán Lozano, I. (2023). Análisis y visualización de los ODS en distintos territorios.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. *European Commission, Ispra*, 15(1), 19-20.

Peña, D. (2013). *Análisis de datos multivariantes*. Cambridge: McGraw-Hill España.

*The Sustainable Development Goals Report 2022*. (2022, May 9). UN Statistics Division. Retrieved July 11, 2023, from <https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf>

(n.d.). CLEAN WATER AND SANITATION: WHY IT MATTERS. Retrieved July 11, 2023, from [https://www.un.org/sustainabledevelopment/wp-content/uploads/2016/08/6\\_Why-It-Matters-2020.pdf](https://www.un.org/sustainabledevelopment/wp-content/uploads/2016/08/6_Why-It-Matters-2020.pdf)

## 8. ANEXO

### Análisis multivariante

Figura A.1.

*Boxplot de cada variable para todos los años del marco temporal*

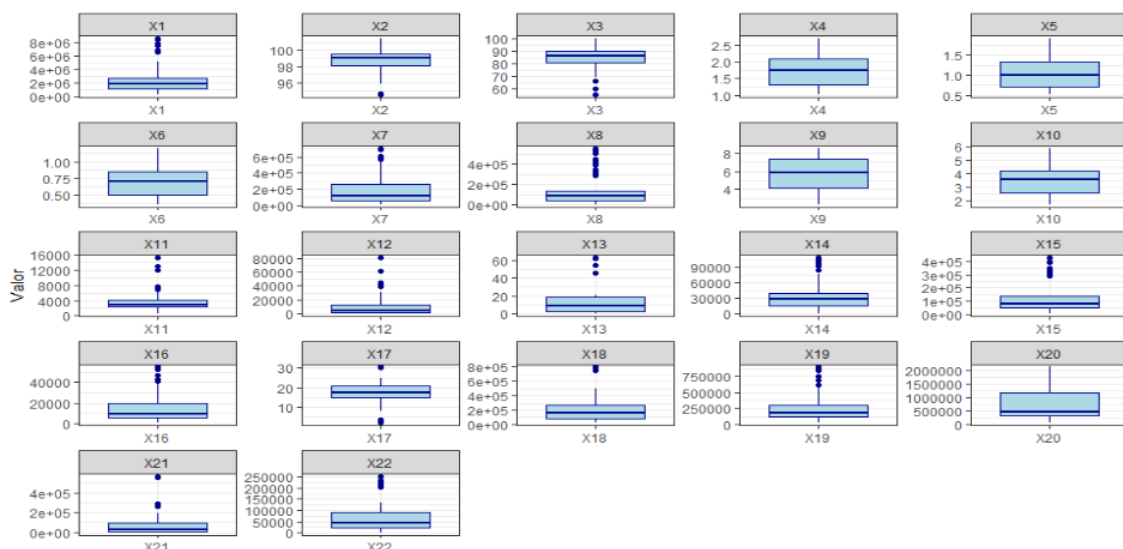


Tabla A.1.

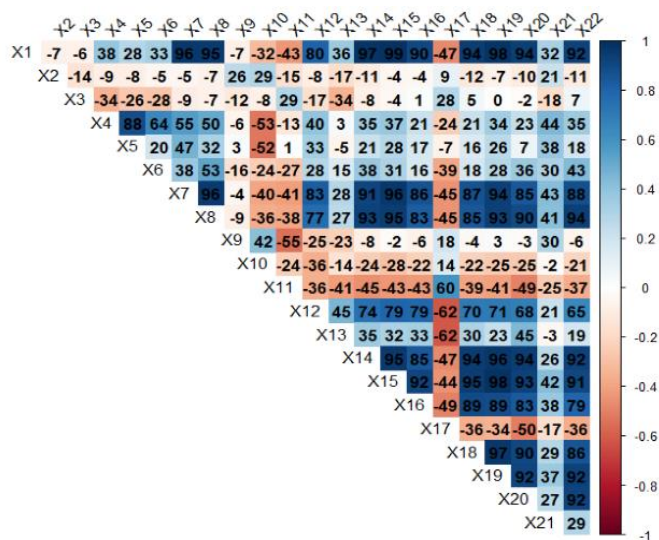
*Vector de medias para cada instante de tiempo*

	Año 2016	Año 2018	Año 2020
<b>X1</b>	2586500,44	2595721,11	2636155,28
<b>X2</b>	97,76	98,50	99,54
<b>X3</b>	84,58	84,09	84,72
<b>X4</b>	1,77	1,76	1,77
<b>X5</b>	1,07	1,05	1,05
<b>X6</b>	0,70	0,72	0,72
<b>X7</b>	207801,06	200677,78	201035,89
<b>X8</b>	138097,06	139003,33	137861,94
<b>X9</b>	5,64	5,77	5,77
<b>X10</b>	3,40	3,49	3,50
<b>X11</b>	3904,83	3562,72	3568,89

X12	12715,17	11803,33	13490,28
X13	15,28	14,89	15,00
X14	35139,11	34958,67	34242,50
X15	127630,67	126178,06	127266,67
X16	15003,00	15977,44	15036,11
X17	18,10	17,12	17,63
X18	226821,22	215319,00	212912,17
X19	275598,89	264983,83	255540,67
X20	719393,11	760254,83	742313,44
X21	75029,78	85227,78	80978,89
X22	65245,83	67247,06	63790,11

**Figura A.2.**

*Representación gráfica de la matriz de correlaciones para todos los años en conjunto antes de las trasformaciones en las variables muy correladas con X1*



**Figura A.2.b**

*Representación gráfica de la matriz de correlaciones para todos los años en conjunto después de las trasformaciones en las variables muy correladas con X1*

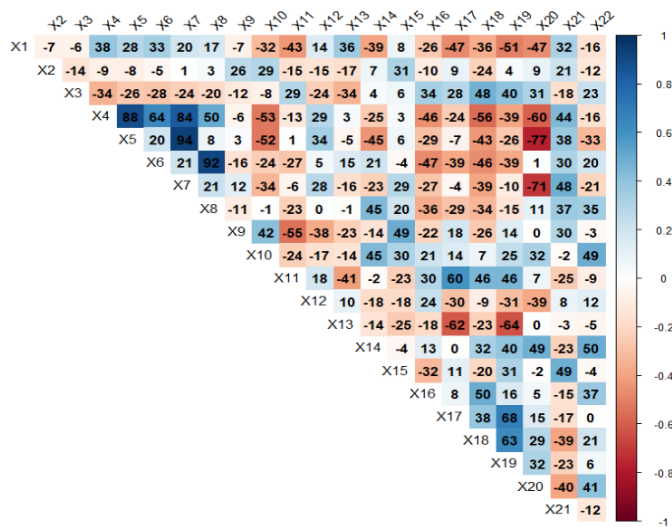


Figura A.3.

Diagrama de dispersión matricial de las variables X1 a X12 y de X13 a X22



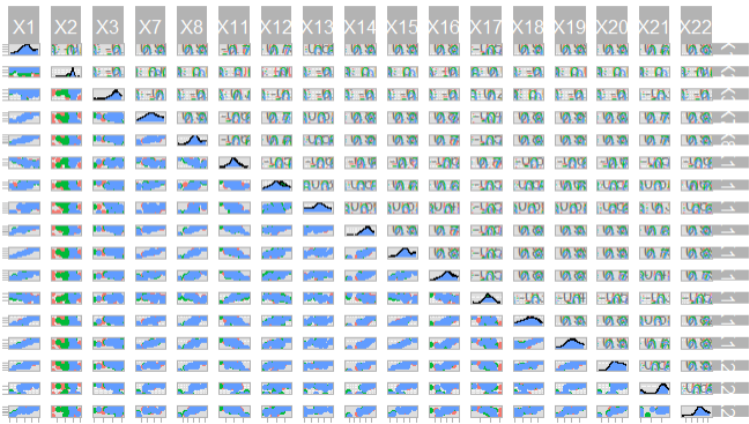
Figura A.4.

Diagrama de dispersión matricial de las variables transformadas antes de ser transformadas



Figura A.5.

Diagrama de dispersión matricial de las variables transformadas después de ser transformadas



PCA: grupo Suministro

Tabla A.2.

Varianzas correspondientes a las variables del grupo Suministro

X3	X5	X7	X9	X11	X12	X13	X14	X15	X16	X17	X18	X19
0,02	0,15	0,11	3,41	0,42	1,19	1,72	0,17	0,02	0,26	32,08	0,19	0,03

Tabla A.3.

Autovectores de cada componente principal del grupo Suministro

	PC1	PC2 <sup>1</sup>	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
X3	0,28	-0,02	-0,06	0,51	-0,04	-0,63	0,21	-0,12	-0,41	0,17	-0,06	0,00	0,01
X5	-0,28	0,45	-0,20	-0,01	0,09	-0,11	0,31	0,18	0,03	-0,11	0,09	-0,16	0,69
X7	-0,26	0,49	-0,11	-0,08	-0,07	-0,25	0,30	0,16	0,24	0,03	-0,14	0,06	-0,65
X9	-0,06	0,17	0,50	0,33	-0,24	0,44	0,20	0,23	-0,29	0,09	-0,10	-0,39	-0,08
X11	0,39	0,22	-0,20	-0,18	0,24	0,01	-0,15	-0,17	0,14	0,34	-0,02	-0,69	-0,07
X12	-0,20	0,12	-0,42	-0,10	-0,46	0,03	-0,46	0,24	-0,39	0,34	0,06	0,05	0,00
X13	-0,29	-0,35	0,14	0,23	0,21	-0,19	-0,06	0,42	0,36	0,48	0,30	-0,09	0,01
X14	0,14	-0,22	0,27	-0,64	-0,30	-0,21	0,37	0,10	-0,05	0,32	-0,19	0,02	0,16
X15	-0,02	0,29	0,44	0,11	-0,36	-0,30	-0,47	-0,22	0,38	0,00	-0,18	0,01	0,21
X16	0,16	-0,15	-0,37	0,30	-0,51	0,28	0,33	-0,19	0,46	0,16	0,00	0,05	0,09

<b>X17</b>	0,35	0,34	0,07	0,09	0,31	0,26	-0,03	0,17	0,04	0,43	-0,23	0,55	0,12
<b>X18</b>	0,39	-0,13	-0,14	0,06	-0,09	-0,12	-0,13	0,68	0,16	-0,40	-0,32	-0,12	0,02
<b>X19</b>	0,41	0,24	0,16	-0,09	-0,19	-0,04	0,03	0,15	0,02	-0,12	0,80	0,11	-0,07

**Tabla A.4**

*Puntuaciones del grupo Suministro*

CCAA	AÑO	PC1	PC2 <sup>1</sup>	PC3	PC4	PC5	PC6	INDICADOR	INDICADOR ESTANDARIZADO
<b>01 Andalucía</b>	2016	0,35	0,25	-0,40	0,51	0,52	0,24	0,11	61,84
<b>02 Aragón</b>	2016	2,15	-0,98	0,58	-0,13	-0,05	1,38	1,17	87,58
<b>03 Asturias, Principado de</b>	2016	0,95	0,51	1,54	-1,26	-2,06	-1,88	0,10	61,70
<b>04 Balears, Illes</b>	2016	1,37	1,59	-0,27	-2,18	-0,62	1,23	-0,06	57,83
<b>05 Canarias</b>	2016	0,18	3,61	0,49	0,22	-0,87	0,41	-0,63	43,98
<b>06 Cantabria</b>	2016	2,13	1,72	2,32	-0,49	1,24	0,74	0,95	82,15
<b>07 Castilla y León</b>	2016	1,35	-0,63	0,74	0,88	-1,54	0,82	0,77	77,79
<b>08 Castilla - La Mancha</b>	2016	0,20	-0,77	0,10	0,44	-0,76	-0,91	0,17	63,19
<b>09 Cataluña</b>	2016	-1,96	0,51	-0,65	0,04	0,42	0,40	-0,87	38,03
<b>10 Comunitat Valenciana</b>	2016	-0,98	1,50	0,29	1,41	-0,64	0,63	-0,50	47,12
<b>11 Extremadura</b>	2016	1,38	0,17	0,43	1,52	0,65	-1,10	0,66	75,30
<b>12 Galicia</b>	2016	-0,20	-0,85	0,85	-0,45	0,75	-0,86	0,23	64,66
<b>13 Madrid, Comunidad de</b>	2016	-3,95	-1,59	-1,21	-0,15	-0,39	1,70	-1,25	28,86
<b>14 Murcia, Región de</b>	2016	-4,09	1,50	0,23	-1,32	-0,69	-1,45	-2,03	9,83
<b>15 Navarra, Comunidad Foral de</b>	2016	1,47	-2,19	-1,08	0,00	-1,11	0,06	0,70	76,30
<b>16 País Vasco</b>	2016	-2,01	-2,50	0,45	-0,02	0,68	0,25	-0,05	57,95
<b>17 Rioja, La</b>	2016	2,79	-1,45	-1,91	-0,71	-0,33	0,07	0,88	80,44
<b>18 Ceuta y Melilla</b>	2016	2,21	2,28	-4,30	0,93	1,98	-0,98	-0,28	52,31
<b>01 Andalucía</b>	2018	-0,12	-0,22	-0,16	0,31	0,25	0,45	0,05	60,46
<b>02 Aragón</b>	2018	1,82	-1,38	0,23	-0,45	0,43	0,42	1,02	83,87
<b>03 Asturias, Principado de</b>	2018	0,04	-0,49	1,12	-0,56	-1,52	-1,31	0,05	60,51
<b>04 Balears, Illes</b>	2018	0,32	1,18	-1,02	-2,19	-0,10	0,90	-0,46	47,89
<b>05 Canarias</b>	2018	-0,19	2,94	0,37	-0,04	-0,50	-0,11	-0,67	42,94
<b>06 Cantabria</b>	2018	1,13	1,76	3,51	-0,46	1,77	1,67	0,91	81,18
<b>07 Castilla y León</b>	2018	1,44	-0,90	0,94	1,15	-0,97	0,50	0,94	81,98
<b>08 Castilla - La Mancha</b>	2018	0,85	-0,34	0,82	0,95	-0,06	-0,82	0,55	72,64
<b>09 Cataluña</b>	2018	-2,36	0,32	-1,00	-0,17	0,29	0,33	-1,07	33,10
<b>10 Comunitat Valenciana</b>	2018	-0,38	1,65	0,99	1,45	-1,41	0,45	-0,26	52,92
<b>11 Extremadura</b>	2018	0,97	0,27	0,52	2,12	1,08	-1,25	0,59	73,47
<b>12 Galicia</b>	2018	-0,50	-2,09	1,30	-0,42	1,59	-0,80	0,53	72,11
<b>13 Madrid, Comunidad de</b>	2018	-3,98	-2,27	-1,57	0,43	-0,58	0,97	-1,20	30,16
<b>14 Murcia, Región de</b>	2018	-5,36	1,94	0,94	-1,33	-0,32	-1,64	-2,44	0,00



15 Navarra, Comunidad Foral de	2018	0,62	-1,59	-1,44	-0,21	-1,23	-0,16	0,16	63,07
16 País Vasco	2018	-1,53	-2,06	-0,06	-0,09	1,26	-0,16	-0,05	58,07
17 Rioja, La	2018	4,15	-1,71	-0,43	-0,96	-0,48	0,38	1,68	100,00
18 Ceuta y Melilla	2018	1,48	2,00	-3,78	0,51	0,47	-0,01	-0,49	47,26
01 Andalucía	2020	-0,40	0,07	-0,22	0,45	-0,06	0,73	-0,11	56,41
02 Aragón	2020	1,43	-0,91	0,72	-1,13	1,20	-0,25	0,82	79,06
03 Asturias, Principado de	2020	1,54	-0,21	1,94	0,22	-0,29	-1,35	0,86	80,14
04 Balears, Illes	2020	-0,93	1,12	-1,38	-3,42	0,08	-0,99	-1,20	29,98
05 Canarias	2020	0,47	2,73	0,18	0,32	0,30	0,25	-0,30	52,00
06 Cantabria	2020	1,09	0,45	2,86	-0,36	0,63	1,27	0,93	81,87
07 Castilla y León	2020	0,94	-1,21	0,68	0,95	-1,35	0,42	0,72	76,75
08 Castilla - La Mancha	2020	0,88	-0,28	0,68	0,20	0,18	-0,61	0,49	71,16
09 Cataluña	2020	-2,84	0,31	-0,54	-0,06	1,43	0,55	-1,04	33,87
10 Comunitat Valenciana	2020	-1,16	0,92	0,32	1,19	-1,65	0,06	-0,58	45,09
11 Extremadura	2020	-0,09	0,08	-0,45	2,17	0,67	-1,60	0,02	59,55
12 Galicia	2020	-0,73	-1,92	1,38	-0,27	1,28	-0,65	0,43	69,66
13 Madrid, Comunidad de	2020	-3,65	-1,78	-1,97	0,34	-0,47	0,80	-1,26	28,51
14 Murcia, Región de	2020	-3,83	1,55	1,18	1,12	0,12	1,11	-1,31	27,42
15 Navarra, Comunidad Foral de	2020	1,03	-1,64	-0,94	-0,07	-1,41	-0,33	0,40	68,80
16 País Vasco	2020	-1,40	-2,33	-0,03	-0,12	1,86	-0,33	0,10	61,58
17 Rioja, La	2020	3,80	-0,74	-0,41	-1,02	-0,04	0,22	1,37	92,56
18 Ceuta y Melilla	2020	2,10	2,12	-3,50	0,24	0,39	0,15	-0,26	52,82

## PCA: grupo Alcantarillado y depuración

Tabla A.5.

*Varianzas correspondientes a las variables del grupo Alcantarillado y depuración*

X6	X8	X10	X20	X22
0,06	0,13	1,16	0,13	0,26

Tabla A.6.

*Autovectores de cada componente principal del grupo Alcantarillado y depuración*

	PC1	PC2	PC3	PC4	PC5
X6	-0,49	-0,50	-0,02	-0,18	-0,69
X8	-0,58	-0,28	-0,18	-0,29	0,69
X10	-0,18	0,63	-0,63	-0,36	-0,21

X20	-0,34	0,46	0,75	-0,32	-0,03
X22	-0,53	0,25	-0,05	0,81	-0,02

**Tabla A.7.**

*Puntuaciones del grupo Alcantarillado y depuración*

	AÑO	PC1	PC2	INDICADOR	INDICADOR ESTANDARIZADO
01 Andalucía	2016	0,14	-0,22	-0,002	51,45
02 Aragón	2016	-1,22	0,30	-0,62	35,13
03 Asturias, Principado de	2016	0,05	1,30	0,54	65,92
04 Balears, Illes	2016	-2,11	-1,13	-1,72	5,85
05 Canarias	2016	3,29	-0,86	1,65	95,33
06 Cantabria	2016	-1,93	0,91	-0,81	30,13
07 Castilla y León	2016	-0,09	1,47	0,53	65,49
08 Castilla - La Mancha	2016	2,13	0,52	1,49	91,12
09 Cataluña	2016	-1,35	-2,64	-1,86	2,19
10 Comunitat Valenciana	2016	-0,82	-1,24	-0,99	25,35
11 Extremadura	2016	1,23	0,94	1,11	81,08
12 Galicia	2016	1,19	1,40	1,27	85,30
13 Madrid, Comunidad de	2016	0,81	-1,38	-0,06	50,04
14 Murcia, Región de	2016	-0,44	-0,24	-0,36	41,94
15 Navarra, Comunidad Foral de	2016	-1,21	1,69	-0,07	49,75
16 País Vasco	2016	-1,45	-0,92	-1,24	18,55
17 Rioja, La	2016	0,25	0,14	0,21	57,05
18 Ceuta y Melilla	2016	3,03	-0,81	1,51	91,71
01 Andalucía	2018	-0,23	-0,30	-0,26	44,68
02 Aragón	2018	-1,20	-0,13	-0,78	30,92
03 Asturias, Principado de	2018	-0,63	0,66	-0,12	48,32
04 Balears, Illes	2018	-2,10	-1,60	-1,90	1,09
05 Canarias	2018	3,44	-0,67	1,82	99,78
06 Cantabria	2018	-1,80	1,06	-0,67	33,73
07 Castilla y León	2018	-0,69	2,12	0,42	62,72
08 Castilla - La Mancha	2018	0,96	1,25	1,07	80,01
09 Cataluña	2018	-1,18	-2,46	-1,68	6,88
10 Comunitat Valenciana	2018	-0,65	-1,03	-0,80	30,36
11 Extremadura	2018	1,30	1,12	1,23	84,19
12 Galicia	2018	0,58	1,10	0,79	72,37

13 Madrid, Comunidad de	2018	1,06	-1,19	0,17	55,95
14 Murcia, Región de	2018	-0,21	-0,69	-0,40	40,83
15 Navarra, Comunidad Foral de	2018	-2,20	2,57	-0,32	43,13
16 País Vasco	2018	-1,12	-1,40	-1,23	18,84
17 Rioja, La	2018	0,11	1,04	0,48	64,24
18 Ceuta y Melilla	2018	2,56	-0,39	1,39	88,48
01 Andalucía	2020	-0,12	-0,47	-0,26	44,59
02 Aragón	2020	-1,28	-0,27	-0,88	28,19
03 Asturias, Principado de	2020	-0,34	1,10	0,23	57,60
04 Balears, Illes	2020	-2,20	-1,54	-1,94	0,00
05 Canarias	2020	3,51	-0,75	1,83	100,00
06 Cantabria	2020	-1,89	0,60	-0,91	27,36
07 Castilla y León	2020	-0,17	1,78	0,60	67,48
08 Castilla - La Mancha	2020	1,95	0,77	1,48	90,90
09 Cataluña	2020	-1,18	-2,20	-1,58	9,55
10 Comunitat Valenciana	2020	-0,47	-1,11	-0,72	32,35
11 Extremadura	2020	0,75	1,00	0,85	74,08
12 Galicia	2020	0,94	1,19	1,04	79,17
13 Madrid, Comunidad de	2020	1,32	-1,36	0,26	58,49
14 Murcia, Región de	2020	-0,29	-0,57	-0,40	40,83
15 Navarra, Comunidad Foral de	2020	-2,10	2,43	-0,31	43,30
16 País Vasco	2020	-1,12	-1,65	-1,33	16,20
17 Rioja, La	2020	0,05	1,02	0,43	63,03
18 Ceuta y Melilla	2020	3,13	-0,28	1,78	98,90

## Construcción del Indicador global

Tabla A.8.

*Varianzas correspondientes a las variables del nuevo conjunto de datos reducido*

Suministro	Alcantarillado y depuración
0,75	1,16